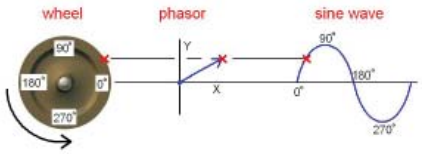
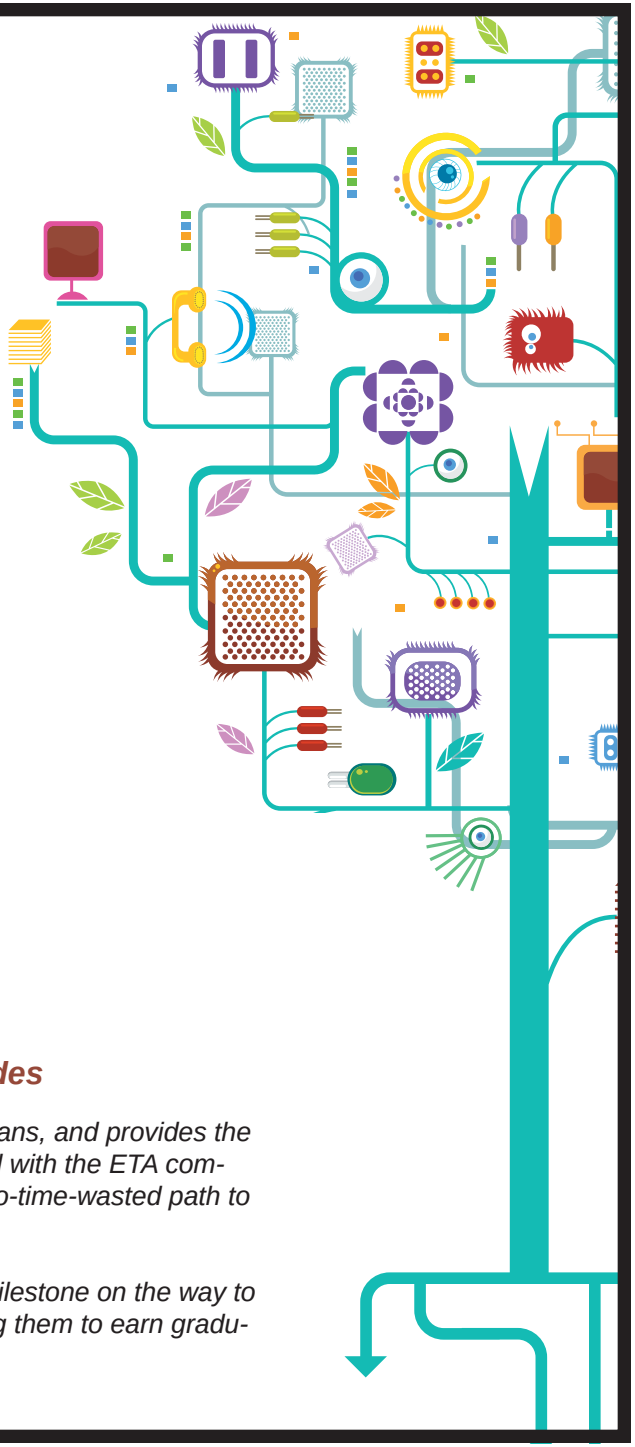


EM2 – AC Basics

by Karl Eilers



EM2 will cover the following concepts:

- AC vs. DC, AC math including frequency, period, wavelength
- Waves and waveforms
- AC test equipment, with emphasis on oscilloscope functions
- Capacitors, inductors, transformers
- RC, LC, RLC circuits, behavior and calculations
- Cabling, characteristic impedance, termination, grounding
- Noise and interference sources and how to handle them

Give Your Students the Power to Succeed with the Online EM-Series Study Guides

The EM-series of study guides give students the knowledge they need to qualify as entry-level technicians, and provides the necessary foundation for further studies in specialized fields. The study guides are closely coordinated with the ETA competencies and exams. Each study guide builds on the one before it. Together they form an efficient, no-time-wasted path to knowledge and certification.

The basic goal of the study guides is the same as the goal of the EM certifications themselves: each milestone on the way to eventual mastery is useful in and of itself, giving students immediately useful information and permitting them to earn graduated certifications commensurate with their growing knowledge.

Study Guide for ETA-International EM1

AC BASICS

by Karl Eilers

Welcome to the second of the ETA EM6 training modules, AC basics. In this module, we will discuss alternating current, AC test equipment, several components that are found only in AC circuits, and the many issues that arise when coupling signals from one piece of test equipment to another. You will need to have mastered the math presented in EM1; in this study guide, chapter 2 presents some additional math that comes into play with AC.

Contents

1. BASIC ALTERNATING CURRENT
2. AC MATH
3. WAVES AND WAVEFORMS
4. TEST EQUIPMENT AND MEASUREMENTS
5. COMPONENTS
6. PRINCIPLES OF CAPACITANCE
7. PRINCIPLES OF INDUCTANCE
8. TRANSFORMER THEORY
9. RC, RL, LC, RLC PRINCIPLES
10. CABLING
11. REVIEW

PRACTICE EXAMS

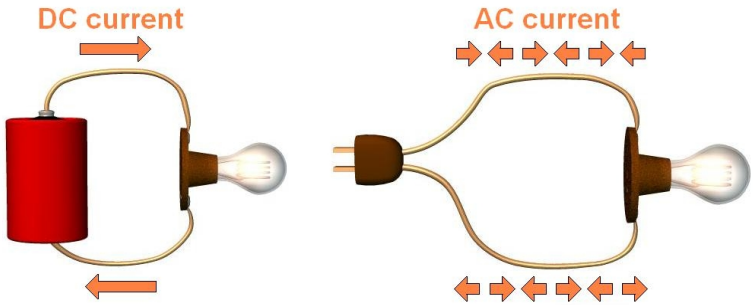
EM2 Chapter 1
BASIC ALTERNATING CURRENT

In this chapter, you will learn:

- *what is meant by "alternating current"*
- *how it differs from direct current*
- *a historical timeline of DC and AC*
- *the schematic representation of an AC voltage source*
- *what "reactance" and "impedance" are, in general terms, and how they relate to resistance*

This study guide builds on the information in EM1. I will assume you've read that study guide and know what current, voltage, resistance and power are, how to do Ohm's and Watt's Laws, how to analyze series, parallel and series-parallel circuits, that you understand basic DC math and how to use a multimeter, and that you have a basic familiarity with magnetism and electromagnetism. Now we're going to move up to the next level and I'm pretty confident you'll come to understand it as well as you do DC.

What's AC, and how does it differ from DC? Direct current flows in only one direction. Alternating current keeps changing direction, back and forth, many times a second.



That is, AC **alternates** back and forth.

A voltage can be a combination of AC and DC, as in a wire carrying a voltage that varies between 4 and 6 volts. These AC and DC aspects can be considered separately, as a 5V DC voltage and a 1V AC voltage. They can even be pulled apart and handled differently..

SIGNAL VS. POWER

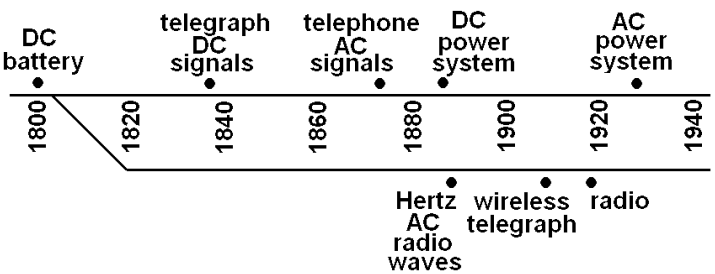
When you send electricity over wires, you're generally trying to do one of two things: moving raw power from one place to another, or moving information from one place to another.

An electrical information stream is called a **signal**. The word originally meant a notification, like a doorbell. Today, the word means *a stream of information that's transported electrically.*

This can be an actual notification, like an alarm bell, or karaoke singer (not every signal is worth sending) or a video picture, or a digital data stream; whether by wires or wirelessly, through the air.

TIME LINE

I thought it might be useful to create a time line, so you could see how AC technology developed. The fact that there was any such thing as AC, or that it could be of any use, did not occur to the earliest researchers.



- ca.1800 – Alessandro Volta invents the battery, which is DC.
- 1838 – Samuel Morse does *not* invent the telegraph. Soon after Volta invented the battery, people were trying to use electricity to send messages. What Morse invented was what we would call software, a code system of long and short pulses that made communication efficient. A switch, called a **key**, is used to create timed interruptions in a DC current that correspond to letters and numbers. The first telegraph message was sent in 1838. The best telegraph operators can still code faster than most people can text.
- 1876 – Alexander Graham Bell and Elisha Grey independently invent the telephone, almost at the same moment. The telegraph sent data; the telephone sent **audio**, which means sound transmitted electrically. Audio is AC, as it consists of positive and negative voltage alternations.
- 1883 – Edison's first DC power plant. By 1900 every major city was wired for DC.
- 1887 – Hertz demonstrates wireless transmission across his laboratory. Today we would call these transmissions **radio waves**, a term derived from “radiation.” Radio waves are always AC.
- 1897 – wireless telegraph. Many researchers had advanced Hertz' work to the point where Guglielmo Marconi thought it practical to start a commercial wireless communication business. Note that this was a telegraph system, even though the telephone had been in use since before Hertz discovered radio. Sending voices by radio waves requires more advanced technology than sending voices by wires. Radio communication was still Morse code in 1912, when the Titanic sank.

- 1901 or 1902 – Marconi transmits radiotelegraph signals across the Atlantic.
- 1918 – **radiotelephone**, “radio” as we know it, with voice and music. The first voice and music broadcast was 1906, but only after World War One ended in 1918 did radio become mass media.
- 1920s – AC power began to replace DC. Part of the problem was conceptual: AC's superiority in a power system wasn't obvious at first. The AC system, invented by Nikola Tesla, scored its first big success when AC generators at Niagara Falls were connected to Buffalo, New York in 1897, just 15 years after Edison's first power plant. Still, the DC system worked pretty well, and many cities were in no rush to convert. The last DC generators supplying power to a building (the New York Library) shut down in November 2007.

So: DC signals came before AC signals, and DC power came before AC power. In very approximate terms (because changes occur gradually):

- Between DC telegraph signals and AC audio signals, 40 years.
- Between the first DC power plant and the beginning of the conversion to AC power, 40 years.
- Between commercialization of wireless telegraph (DC signals carried on AC radio waves) and radio as we know it (AC signals carried on AC radio waves), 20 years.

AC IS MORE COMPLEX THAN DC

The reason AC took longer is because of its greater complexity; technology had to advance to the point where it was practical to generate and handle AC.

DC is comparatively simple. Voltage is present or it isn't, current flows or it doesn't, current paths are closed or open. If you state the voltage, the current, the resistance it flows through and the resulting power dissipation, you've described everything about it. AC, on the other hand, is constantly reversing direction. The fact that it's always changing means you now have to consider *how* it's changing. With AC there's a lot more to describe.

DC could be compared to a wall that's painted a single color: if you name the color you've said everything there is to say. But if it's painted two colors, there's more to describe than just the colors:

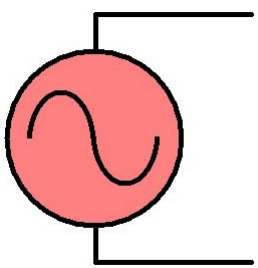
- What part of the wall is one color, and what part is the other?
- Is the pattern regular or random?
- Are the colors kept separate, or are they mixed in some places?
- Is the ratio of the colors 50:50, or does one color predominate?

Just by virtue of the fact that we've gone from one color to two, a bunch of new questions arise. Likewise, since AC alternates, we have to describe how often, in what pattern, and so on.

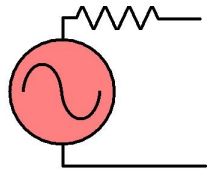
VOLTAGE ALTERNATIONS

Just as AC current reverses direction, an AC voltage reverses polarity, alternately going positive and negative. We will often need to look separately at the two parts of a voltage pattern, so let's name them. We'll call them the **positive alternation** and **negative alternation** respectively.

AC VOLTAGE SOURCE SYMBOL



We used the symbol for a battery to represent a generic DC voltage source. The generic AC-source symbol is a circle with what's called a **sine wave** in it. You've probably seen that curvy shape before, as a symbol for an AC voltage. I'll have a lot to say about sine waves in chapter 4, they're very important. This symbol can be used to represent any AC source – a power line, an MP3 player, a video camera. We think of this kind of source as having zero internal resistance – that is, it's an ideal, **stiff** voltage source. A real-world voltage source has some internal resistance. Where this resistance will have a significant effect on circuit behavior, it may be represented in a diagram by a perfect voltage source in series with a resistor.



REACTANCE AND IMPEDANCE

When working with alternating voltage and current, there's an odd phenomenon we run into all the time: voltage and current are often out of step with each other. You recall Ohm's Law from EM1:

$$V = I \times R$$

This implies that if voltage changes, current will change with it, instantly. This is true of AC as well as DC, as long as the resistance is just resistance and nothing else.

But in this study guide, we're also going to talk about **capacitors** and **inductors**. Like resistors, these components oppose current flow, and their opposition is measured in Ohms. Unlike resistors, they do not dissipate power; they store and release it. And unlike resistors, current flow *through* them is out of step with the voltage *across* them.

Since they exhibit opposition to current flow, but the opposition is not resistance, it has been given another name: **reactance**.

In many circuits, capacitors and/or inductors are combined with resistors. This means the opposition to current flow is neither pure resistance nor pure reactance. So we need yet another term to describe this mixed opposition.

The general term for opposition to current flow is **impedance**. In a way, it's just another word that says the same thing: something that resists, impedes, works against current flow. The reason for having a separate word is that it gives us a name that covers all types of opposition to current flow, just as “people” covers Lithuanians, Albanians, Ukrainians and everybody else.

When you see the word “impedance,” think: maybe this is a pure resistance, maybe not. We use the word in reference to microphones, speakers, antennas, cables, equipment inputs and outputs and circuits inside the equipment among other things.

AC MATH

Dealing with reactance and impedance, and other aspects of current that keeps switching direction, requires some math we didn't cover in EM1. If you're doing this study guide without first doing EM1, you need to download its math chapter and study it, because the math you use with DC is the foundation of the math you will use with AC.

Additional AC-related math in this study guide:

- expanding graphs to handle negative numbers
- mathematically representing time and change
- trigonometry, the mathematics of angles
- measurement scales that are proportional rather than linear, when we want a graph to show many orders of magnitude

Since we'll need that math almost immediately, it comes next, in chapter 2. Then, in chapter 3, we'll examine all those new issues that cropped up when we went from DC to AC. They are complications, but they are also opportunities. You can do a lot more with AC than with DC.

End of EM2 Chapter 1

EM2 Chapter 2
AC MATH

Objectives of this chapter:

- *know how to graph positive and negative numbers*
- *calculate trigonometric ratios*
- *make calculations involving circular motion*
- *understand proportional and logarithmic scales*

Mathematically, the main difference between AC and DC is that AC voltage (and current, and instantaneous power) is always changing. Of course, strictly speaking, DC changes too – batteries run down, for example. But AC is based on change, often billions of changes per second. So, if you ask, “What is the AC voltage at this point in the circuit?” the proper answer is, “At what moment?” The AC voltage at a power outlet spends as much time being negative as positive, so the average voltage is zero; but if you've ever gotten a shock, you know the effective voltage is *not* zero. So, special ways of measuring AC have been developed. In chapter 3, we'll study the various characteristics of AC and how they are measured. In this chapter we'll deal with the math that those measurements are based on.

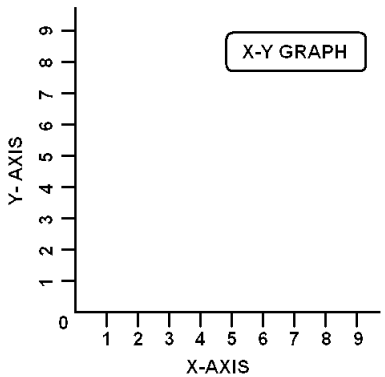
In this chapter, we'll be dealing with angles a lot, which means we'll be dealing with something called **trignonometric ratios**, so you need a scientific calculator. The cheapest one you can get will be adequate, though if you get one that can translate between decimal numbers and octal and hexadecimal, you won't have to buy another one when we get to the EM4 study guide. Until you get a calculator, you can probably find one on your computer which can be set to scientific mode. Open the calculator in a separate window, then minimize it but don't close it; you will want to be able to bring it up often. When we get to the part on trigonometry, you may want to resize this tutorial window so you can show the calculator in a window next to it.

UPPER AND LOWER CASE LETTERS

There's a convention that DC quantities are represented by upper-case letters **V**, **I**, **R** and **P**, while AC quantities are represented by lower-case letters **v**, **i**, **r** and **p**. The lower-case letters are often italicized as in ***v, i, r, p***. You'll find this mostly in equations. You won't find it in the nuts-and-bolts everyday world, e.g. building power is 117VAC, not 117vac. There's no hard-and-fast rule. In these study guides, I'll be using both upper and lower case, depending on what fits the particular situation. Just so you know.

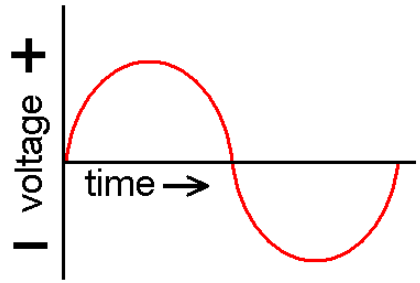
TWO- AND FOUR-QUADRANT GRAPHS

When we're dealing with AC, we use graphs a lot more because we need to show how a quantity changes with time. The first thing we need to do is expand our concept of X-Y graphs. With graphs like those we used in EM1, where the origin is at the lower left corner, we can only display functions where the values of X and Y are always positive. We know there are such things as negative numbers, and when we start talking about AC, we immediately need them.



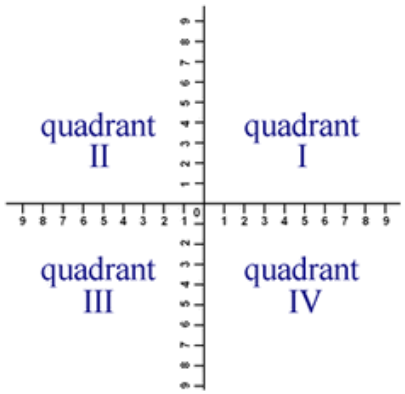
Is there such a thing as a negative voltage? Of course there is, an AC voltage continuously alternates between positive and negative. As we saw in EM1, time is usually plotted on the X-axis and voltage on the Y-axis. With AC, voltage at a given moment can be positive, negative, or zero. So we need to expand our graph to show negative voltages. We do this by extending the Y-axis downward from the origin, so it goes negative as well as positive. Now the X-axis is in the middle of the graph instead of at the bottom.

You've seen a sine wave, which we're going to discuss in considerable detail later on. Here's what it looks like on the graph.



Is there such a thing as negative time? Maybe the Big Bang could be considered “zero” time, in which case all time since then is positive time. But that's not very useful in electronics. Most often, we want to latch onto a particular moment in the continuous flow of time, a moment when something important happens, and place it at the origin of a graph. That becomes “time zero.” In that case, as far as we're concerned, any time before that is negative time. NASA identifies the moment when a rocket engine ignites as zero, and counts down “T-minus” time before that. In electronics, the circuit condition at a specific moment often depends on the previous condition. We'll also deal with circles centered on the origin. So we'll need to extend the X-axis to the left of the origin.

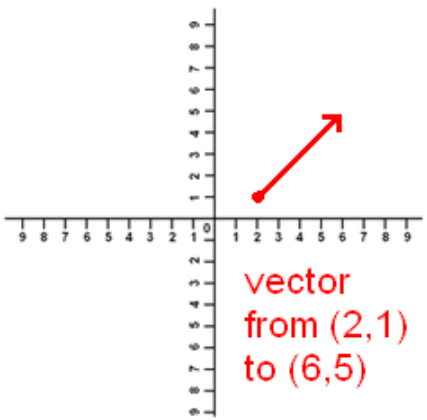
This gives us a graph divided into four parts. These are called **quadrants** (quad because there are four) and mathematicians identify them with Roman numerals. The quadrant we already know, where values of X and Y are both positive, is quadrant I.



Then, going counterclockwise: quadrant II shows points with negative X and positive Y; quadrant III shows points where both X and Y are negative, and quadrant IV shows points with positive X and negative Y. We make graphs using whichever quadrants are needed to show what we want to show.

CIRCLES, VECTORS, PHASORS AND TRIGONOMETRY

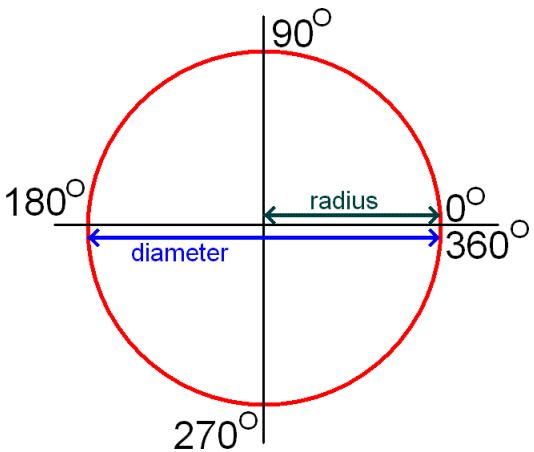
A **vector** is an arrow with a specific length and direction that shows movement or displacement, as in the illustration at right. The vector begins at coordinates 2, 1 and ends at 6,5. It could just as easily pointed into or out of the plane of the graph; that is, a vector can exist in 3-dimensional space.



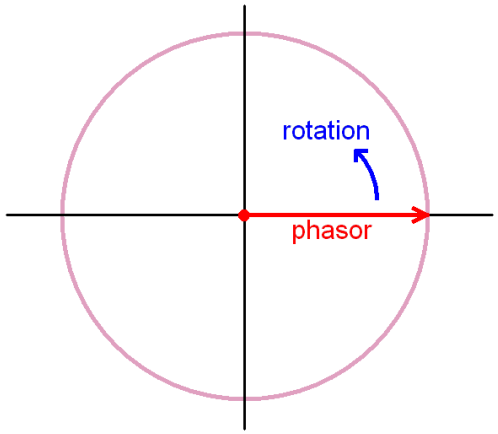
A **phasor** is a special kind of vector, an arrow whose "tail" stays in one place and whose point rotates counterclockwise at a constant rate.

As you know, a circle is normally divided into 360 degrees. To draw a phasor diagram, we first draw a circle with "zero degrees" placed to the right on the circle. This allows us to put the circle on an X-Y graph, with the circle centered on the origin and a point on the positive X-axis that corresponds to the size of the circle.

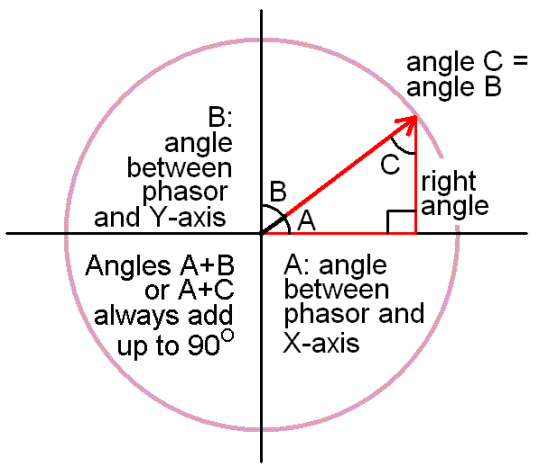
Then, from this zero-degree point, the 360 degrees go counterclockwise around the circle, so 90 degrees is up, 180 degrees is left, 270 degrees is down, and by the time we reach 360 degrees, we're back at the X-axis. The end of one trip around the circle is the start of the next trip; 360 degrees is the same point as zero degrees.



The distance from the center of the circle to the edge, that is, from the origin to the point on the X-axis that intersects the circle, is the circle's **radius**. A circle's **diameter** is twice its radius. Yeah, you knew that.



Next we draw the phasor, an arrow that starts at the origin and points to the right, so it ends on the X-axis at the edge of the circle.

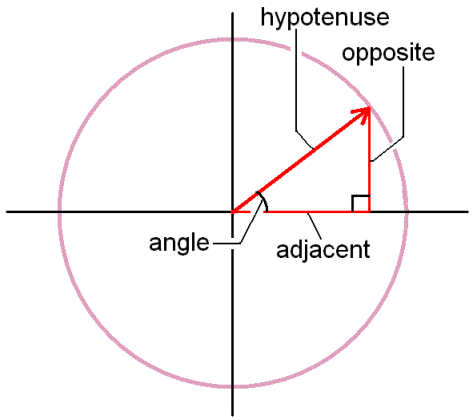


Suppose we rotate the phasor counter-clockwise, so the arrow's tail stays at the origin but the point follows the circle. At each position of the phasor, it forms an angle with the X-axis ("A" in the illustration at left) and a **complementary angle** with the Y-axis ("B" in the illustration). By "complementary" I mean that, as the X-angle gets larger, the Y-angle gets smaller; the two angles always add up to 90 degrees.

The position of the tip in X-Y space depends on the length of the phasor and the angle. Suppose the phasor is 3 "units" long, the units being inches, volts or whatever. If it's at a 35 degree angle, you could describe the position of the tip by saying it's at a 35 degree angle from the origin and 3 units away from it. You could enclose the two quantities in parentheses to denote that you're describing a point: (3 inches, 35 degrees). As the phasor rotates, the tip of the arrow traces a circle with a 3-inch radius or 6-inch diameter.

If you set the phasor to 35 degrees and draw a line from the tip straight down to the X axis (the vertical red line in the illustration at left), you've just made a triangle consisting of the X axis, the phasor, and the new line you drew. This is a **right triangle** – it has a right angle in it, because the line you drew is at a 90 degree angle from the X axis – which permits us to use known mathematical formulas to find the phasor tip's X coordinate. We can also find the length of the line you drew, which is cool because the length of that line is the height of the tip, which is its Y coordinate. So, these formulas let us convert the phasor length and angle into x,y coordinates.

It's often more useful to describe the position of the phasor's tip as a pair of x and y coordinates, because when we have just numbers instead of numbers and angles, the phasor's position, and the distance it moves, are easy to calculate. When the phasor is at zero degrees, pointing to the right, figuring out the tip's x,y coordinates is a no-brainer. X is 3, Y is 0. So (x,y) = (3,0). Likewise, when it's straight up, at 90 degrees, the coordinates are (0,3). Straight down is (0,-3) and to the left is (-3,0). But what about 35 degrees, or some other angle? This is where we get into **trigonometry**, which involves the study of triangles.



We need to name the lines that form the triangle, so we don't have to keep saying "the line you drew." If you look at the angle between the X-axis and the phasor, you see that the triangle has a horizontal line next to the angle, which is called the **adjacent**, because it's adjacent to the angle, and a vertical line across from or opposite to the angle, called the **opposite.** The slanted line that joins them is called the **hypotenuse**, which sounds to me like an animal created by an accident in a genetics lab.

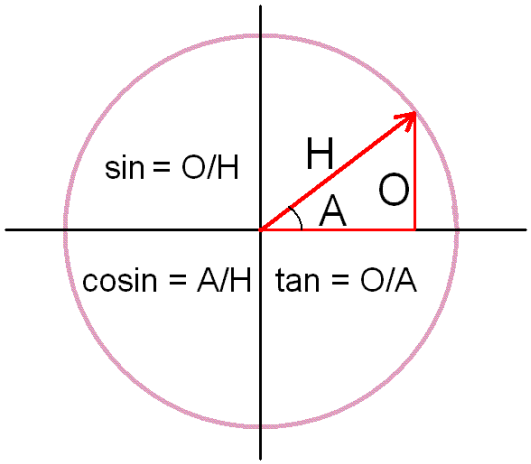
The hypotenuse is always the longest of the three. Trust me. Or don't; try to draw a right triangle where the opposite or adjacent is longer than the hypotenuse. Remember, we're talking about *right* triangles – one of the angles has to be 90 degrees.

Trigonometric Ratios

There is a specific ratio between the length of one side of a right triangle and either of the other sides. These ratios depend on the angles in the triangle, and we use them to convert the length and angle of the hypotenuse – that's the phasor, remember – into x,y coordinates. Or, sometimes, we have x and y coordinates and want to find the length of the hypotenuse. The ratios are so important, and so frequently used, that they've been given names. You'll need to memorize them.

When we studied math in EM1, we saw that a ratio states the proportion between two numbers, which are separated by a colon, as in 3:2. If you restate this as a fraction, $^{3}/_{2}$, and then go ahead and perform the implied division, you get a single decimal number: $^{3}/_{2} = 1.5$. The trignonometric ratios we're going to talk about are decimal numbers of this kind.

The length of the opposite divided by the hypotenuse is called the **sine**, the adjacent divided by the hypotenuse is the **cosine**, and the opposite divided by the adjacent is the **tangent**. Mathematicians abbreviate these quantities **sin**, **cosin** or **cos**, and **tan**. Print out the illustration (it's duplicated in the printable review chapter at the end of this study guide) and keep it handy until you've memorized the names.



Since the hypotenuse is always longest, the sine and cosine will always be less than 1. The tangent can be more or less than 1. At zero and 180 degrees the tangent is zero. At 90 and 270 degrees it's infinity.

There is no easy way to find these ratios with paper and pencil. Math books used to have pages of sines, cosines and tangents, but who wants to carry a math book around on the job? Fortunately, scientific calculators have come to the rescue. You should memorize a few representative sines, cosines and tangents (for example, the sine of a 30 degree angle is ½, and the tangent of a 45 degree angle is 1) so if you punch the numbers in wrong, you'll know the answer is way off from what it should be.

THE PYTHAGOREN THEOREM

In a right triangle, the squares of the opposite and adjacent, added together, equal the square of the hypotenuse.

$$O^2 + A^2 = H^2$$

This is known as the **Pythagorian theorem**. It was discovered 1,000 years before Pythagoras, but he was the one who gave us a mathematical proof, about 1,500 years ago, which is why it's named after him. Various arrangements of the formula are at right. It can't be said too often, so here's another reminder: these formulas only apply to right triangles.

$$H = \sqrt{O^2 + A^2}$$

$$O = \frac{H^2}{A^2}$$

$$A = \frac{H^2}{O^2}$$

One familiar triangle, for instance, is the 3-4-5 triangle, where the short sides are 3 and 4 inches long, respectively, and the hypotenuse is 5 inches. Check it out: 3^2 is 9, and 4^2 is 16. 9 + 16 = 25, which is 5^2.

Just in passing, this will come in handy – the sum of all the inside angles of a triangle is always 180 degrees. In a right triangle, one corner takes up 90 degrees of that, which means that the sum of the other two is always 90 degrees.

TRIG EXERCISES

If you and your scientific calculator don't know each other's secrets by now, you're about to find them out.

1. Opposite: 3 inches. Adjacent: 4 inches. How long is the hypotenuse?
2. Opposite: 5.7 inches. Hypotenuse: 8.1 inches. Which quantity – sin, cosin, tan – describes this ratio? What is the number? What is the angle?
3. Sin = 0.82. If the adjacent is 3 inches, how long is the opposite?
4. The top of a ramp is 30 feet from where it touches the ground and 5 feet above ground level. What is the angle of the ramp? How long is it?

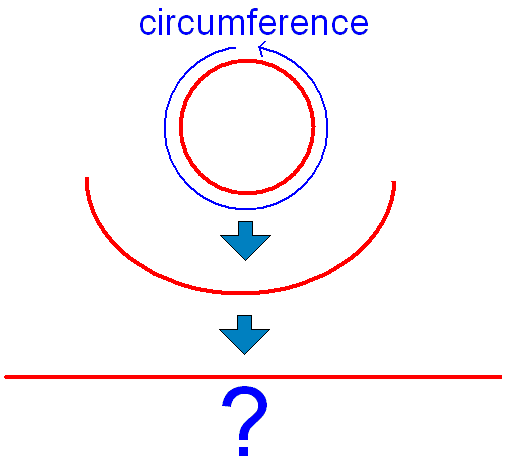
Answers:

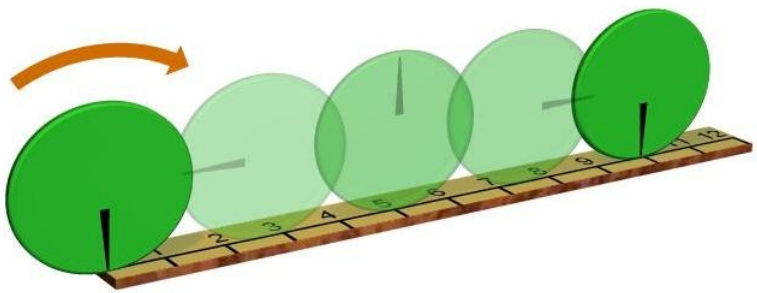
1. I already gave you this one, **5 inches**. 3^2 is 9, 4^2 is 16, 9 + 16 is 25. $25^{1/2}$ is 5. Yeah, I'm being a jerk, forcing you to work with fractional exponents. If you get good at that, techs who aren't will burn incense in your honor.
2. Opposite over Hypotenuse is the **sine**. 5.7 / 8.1 is about **0.7**. A sine of 0.7 corresponds to about **44.7 degrees**. Just as a point of reference, 45 degrees has a sine of 0.707, the square root of two.
3. To the nearest 2 digits, the angle that has a sine of 0.82 is 55 degrees. The tangent (op/adj) of 55 degrees is 1.4. This means the opposite is 1.4 times as long as the adjacent. 1.4 x 3 = **4.2 inches**.
4. Opposite = 5ft, adjacent = 30ft. Tan = 5/30 = 1/6 = about 0.167. That corresponds to a 9.6 degree angle. The cosine of a 9.6 degree angle, adjacent/hypotenuse, is about 0.99. We know the adjacent, it's 30 feet, so to find the length of the ramp we find the reciprocal of 0.99. 1/0.99 = 1.014. Times 30, that's **30.4 feet**.

CIRCUMFERENCE, RADIANS, ANGULAR VELOCITY

We often need to know how far the tip of a phasor travels in its path around the circle – that is, the length of the line that defines the outside of the circle. If we broke it and straightened it out, how long would it be?

The word that defines this length is **circumference**, and it's a royal pain to measure. If you had a circular disc in your hand, it wouldn't be so bad. You'd lay a ruler on the table, put a mark on the edge of the disc, put the disc on the ruler so the mark lines up with the ruler's end, and roll the disc along the ruler until the mark came around again.





But what about a circle that's just a line on a paper or a shape on a computer display?

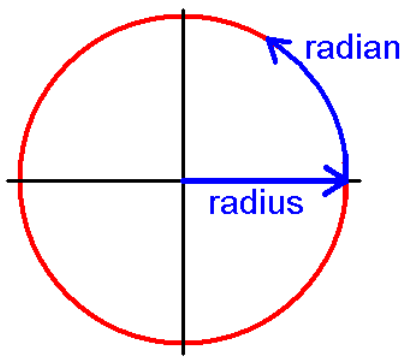
Fortunately, there's a mathematical relationship between the diameter of a circle and the circumference. Unfortunately, it's not an integer. In fact, it's irrational. The circumference of a circle is 3.14159265358979... times the diameter, and like any irrational number with an ounce of self-respect, its digits go on forever. Generally, in electronics, we can truncate this to 3.14 and be close enough. This ratio is called **pi**, long "i" as in "pie," and it's represented by the Greek letter π:

$$\pi = 3.14$$

Since the diameter is 2 times the radius, the circumference is 2π times the radius:

$$C=2\pi R$$

and in this case C is circumference, not capacitance, and R is radius, not resistance.



This can also be stated as: the circumference is 2π **radians**. What's a radian? Imagine that you took a stick the length of the radius and bent it around the edge of the circle, so it followed the circle for the length of a radius. That's a radian. It would take 6 sticks that length, plus about 28% of another stick, to go all the way around the circle. Yup, 2π radians.

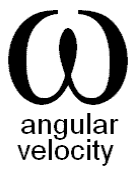
Angular Velocity

The reason for bothering with radians is that we often have to know not only where the tip of the phasor is, but also how fast it's moving around the circle – not just position, but change in position per unit of time. In plain English, speed.

Suppose the phasor is rotating, with the tip moving around the circle in a counter-clockwise direction and the other end anchored at the origin. Suppose it makes one complete revolution per second. How fast is the tip moving ? Well, that should be easy. If the phasor is making one revolution per second, and the circumference is 2π radians, then the tip is moving a distance of 2π radians per second. So, let's say the phasor is 2 inches long, meaning the radius and radian are 2 inches long. That means the tip is moving 2π times 2 inches per second:

angular velocity = 2in x 2π per second
angular velocity = 2in x 6.28 per second
angular velocity = 12.56in/s

Now suppose the phasor is making 60 revolutions per second. That means the tip is moving 2(2π) inches per second times sixty, or 120(2π) inches per second. We can work this out all the way: 120 times 2 is of course 240, times 3.14 = 753.6 ips (inches per second).



The speed of the tip as it travels around the circle is known as **angular velocity**, it's measured in **radians per second** and its symbol is the lower-case Greek Omega, **ω**. Why'd they choose that? Well, look at it – reminds you of a phasor going in circles, doesn't it.

Is this angular velocity thing of any use? Yes, you'll run across it often. But what's more useful is to divide angular velocity by 2π to get **frequency**, which is the number of times per second the phasor makes a complete revolution. You won't actually be doing that operation by itself very often, but it's part of many formulas, which have a 2π term in them.

In fact, you should memorize that 2π is about 6.28, or 6.3 if you want to be lazy – and my friends would tell you I always want to be lazy, I'm as dependable as death and taxes in that way. Using 6.3 involves only about 0.3% error. I'm going to use 6.3 for 2π throughout the following chapters and remaining study guides.

ELECTRONIC MATH SYMBOLS

This is probably a good time to deal with something that scares a lot of people. A beginning student often looks at a formula full of odd symbols and thinks, “How am I supposed to make heads or tails of that? It might as well be Greek.”

There is no reason to be afraid of strange symbols. They really exist to make a technician's life easier. “Yeah, right,” I hear you say, “and you're really the king of North Dakota.” Okay then: suppose you're working with a formula for angular velocity. Mathematical progressions sometimes involve a dozen steps. Take a pen, right now, and write “angular velocity” out in longhand twelve times. Then write “ω” twelve times. Which would you prefer? Thought so.

That's all those mathematical symbols are – they're a way to indicate concepts, relationships and operations without going to the bother of writing them out in words. You're using mathematical symbols already. When you add, you don't write “plus,” you write “+,” and you've known how to do it since you were six. This is just more of the same. Mathematicians are as lazy as the rest of us. If they do something a certain way, you can bet it's the simplest way to do it.

And, by the way, most of those symbols *are* Greek. Nobody uses the ancient Greek alphabet any more, not even Greeks, and it's easier for lazy mathematicians to borrow those symbols than make up new ones.

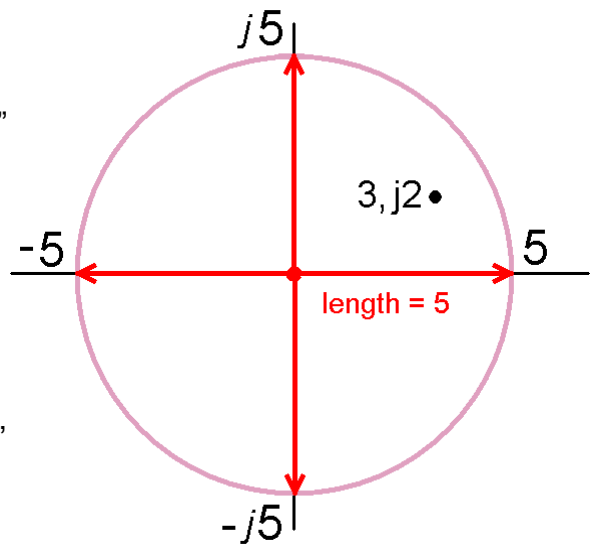
So don't let this stuff scare you. The main reason most people can't handle math is because they think they can't. They see Greek symbols and their eyes glaze over. But there's nothing magic here; most of the stuff we're going to deal with isn't even very complicated.

REAL AND IMAGINARY NUMBERS

Does a phasor going to the right on the X-axis from the origin represent a positive or negative number? Positive, right. Does a phasor going to the left from the origin represent a positive or negative number? Negative, sure thing. But what about a phasor pointing straight up? It's neither positive nor negative, but it isn't zero; it doesn't begin and end at the origin. So what do we call the number it represents?

Mathematicians call numbers that fall on the X-axis **real numbers**. They call numbers that fall on the Y-axis **imaginary numbers**. Dunno why, imaginary numbers are perfectly real. Dumb name if you ask me. But that's what they're called.

An imaginary number is usually indicated by a lower-case italic "*i*" in front of it, like *i5*, indicating +5 on the Y-axis. However, electronics people have already decided that "*i*" means AC current, so for imaginary numbers we just shift over one letter, using "***j***" to indicate an imaginary number.



A point on an X-Y graph that is somewhere besides one of the axes can be represented, as you know, by a pair of numbers like (3,2), indicating a point that's in quadrant I, three units to the right and two units high. This is often written as **3, *j*2**. This is called a **complex number**. A bit later on, we're going to start talking a lot about **phase shift**, which describes the relationship between two sine waves that aren't precisely in step. That involves complex numbers.

PHASE

A phasor is said to have **phase**, which refers to the position of the phasor's tip in degrees; when it's pointing to the right along the X-axis, that's defined as zero degrees. The concept of phase is especially useful when we have several phasors whose tips are at various points around the circle. For example, electrical power comes from the generating plant with three phases on three wires, which can be mathematically represented by three phasors each displaced 120^{O} from the others, rotating counterclockwise together at a constant speed. In fact, "phasor" means "phase vector."

PROPORTIONAL SCALES

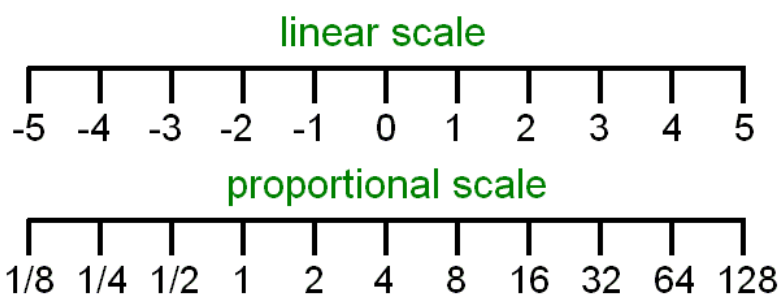
On a ruler, the space occupied by an increment indicates its significance. On a normal linear ruler the distance between each pair of marks is equal to every other pair. We're so used to linear rulers that it seldom occurs to us that there's any other way to measure.

But in our daily lives, we live in a *proportional* reality. If a $1 bag of French fries goes up to $2 overnight, you're irked. But if a $30,001 car goes to $30,002, you don't care. It's the same buck, but one buck in $30,000 hardly matters. The difference between a 3-year-old and a 5-year-old is huge, but when those two people get to be 53 and 55 years old, the difference will be insignificant.

In electronics too there are many cases where a number's significance depends on context. Or, we may have to graph quantities that vary over may orders of magnitude. That's when we use **proportional scales**.

Would you say the increments between 1, 2, 3, 4 and 5 are equal? Of course they are. And, on a linear scale, each increment occupies the same amount of space.

Would you say the increments between 1, 2, 4, 8 and 16 are equal? Yes, in a different sense: they are proportionally equal. On a proportional scale, each proportional increment occupies the same space.



A proportional scale has no zero. Going to the left from 1 on a proportional number line, the increments are ½, ¼, $^1/_8$ etc.

A proportional number line shows what a number means in its context. The space an increment occupies depends on its significance; a small increment occupies a significant space in a neighborhood of small increments, but a much smaller space in a neighborhood of larger increments.

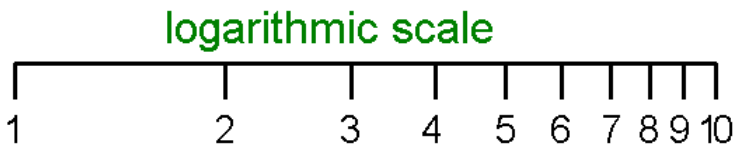
For example, on a linear scale, the distance between 1 and 2 is the same as the distance between 11 and 12; one space in either case. But on a proportional scale, the distance between 1 and 2 is large, while the distance between 11 and 12 is small. This is because the difference between 1 and 2 is 100%, but between 11 and 12 is only about 9%.

On the other hand, the proportional relationship between 1 and 2 is the same as between 10 and 20, so on a proportional scale, those two increments occupy equal amounts of space.

THE LOGARITHMIC SCALE

The problem with a proportional scale is that it wants to use numbers like 2, 4, 8, 16 while we think in terms of 1 to 10.

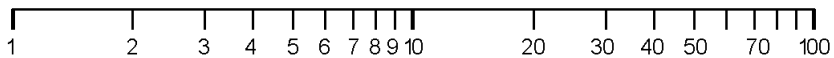
The way we solve this problem is to insert the extra numbers between the proportional scale's numbers, taking care to position them proportionately.



This is called a **logarithmic** or **log** scale. I'm not even going to start a discussion about what logarithms are, but they have to do with relative proportions among decimal numbers.

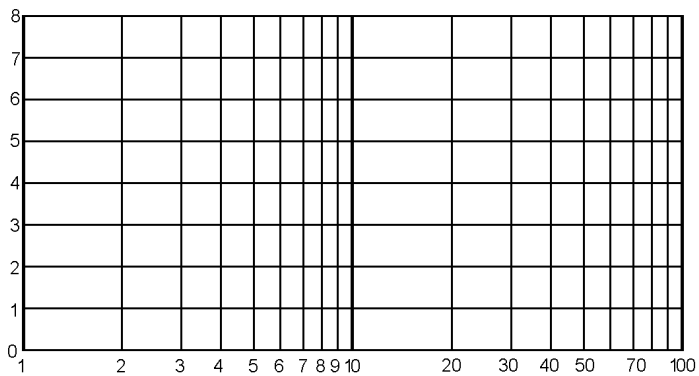
On a logarithmic scale, integers near 1 are spaced apart, while integers near 10 are crunched together. If you take a pair of dividers and measure the increments, you will see that proportional spacing has been preserved; the distance between 1 and 2 is the same as between 2 and 4 and between 4 and 8. These are the same numbers in the same places they would have been on a normal proportional scale; we've just fitted the rest of the decimal numbers around them.

If we want, we can do this again. We can extend the number line, going from 10 to 100, and again, going from 100 to 1,000 and on as far as we like. Each time we do this, we are expanding the scale by an order of magnitude. Every 1-to-10 span is called a **cycle**. This could lead to some confusion; as you'll see in the following chapters the word “cycle” usually means something entirely different to an electronics person. You just have to keep the two definitions separate in your mind.



The illustration below left shows two cycles. Note that in the second cycle, the increment between 10 and 11 doesn't even show up, because an increment of 1 is too insignificant. Note also that the increment between 10 and 20 is the same size as the increment between 1 and 2.

If the X-axis of a graph is a log scale, does that mean the Y-axis will be a log scale? Not necessarily. The illustration below shows a **log-linear graph**, with a 2-cycle log scale on

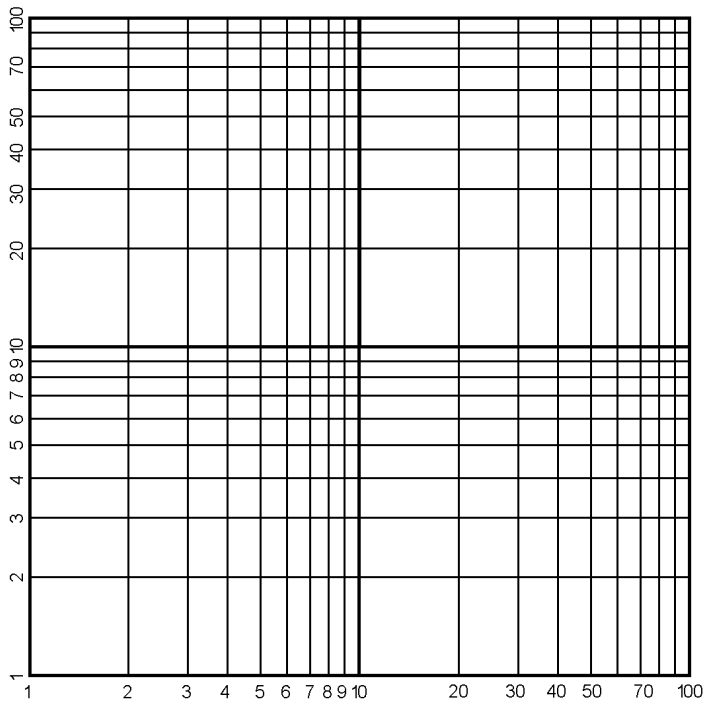


the X-axis and a linear scale on the Y-axis. This is the same as any other graph that plots Y as a function of X, except that the X-quantity is shown logarithmically.

A log-linear graph would be used to show a quantity that spans two orders of magnitude, but exhibits a function that varies within a reasonably narrow range. It is not necessary for a cycle to start or end at 1, 10, 100 etc. It only needs to have as many graduations as needed to show the range you're trying to show. For example, the audio spectrum is often shown on a 3-cycle graph beginning at 20Hz and ending at 20Khz.

The **log-log graph** to the right would be used when both X and Y quantities span a two-orders-of-magnitude range.

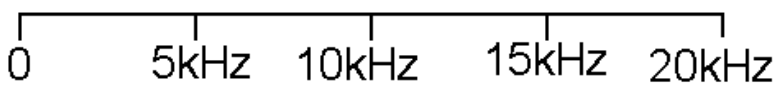
In some cases, a logarithmic characteristic is built into the way something is measured. The most common example is the **decibel**, which techs use a lot and which we'll be studying in EM3. This is a logarithmic quantity where the numbers represent orders of magnitude: 10, 20, 30 and 40dB represent 10, 100, 1,000 and 10,000 units. Since the proportionality is built into the way decibels are measured, we can use a graph where the Y-axis is linear. But the frequency span might be 20-20,000 Hz, which would require the X-axis to be logarithmic. So, the proper graph would be log-linear.



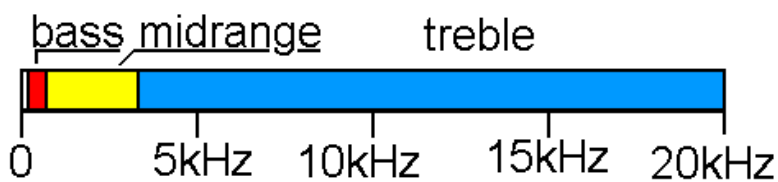
CHOICE OF SCALE

To some extent, the choice of scale type depends on the range of numbers, e.g. to graph a quantity that varies over several orders of magnitude, you'd tend to choose a log scale.

But the choice should really be governed by what the numbers mean in the real world. For example, if we take the range of human hearing to be 20Hz to 20kHz, we could put that on a linear scale like this:



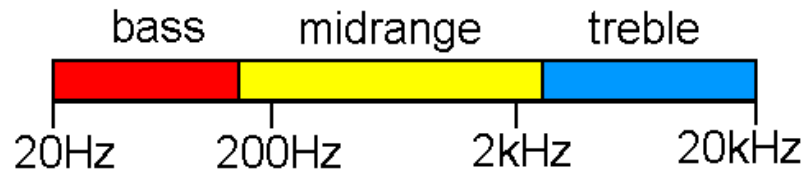
That's a perfectly legitimate scale, but it presents a problem. Our ears hear frequencies logarithmically. Audio is commonly divided into three regions – Bass, Midrange, Treble – because these are practical real-world descriptions of what people hear, subjectively. But look what happens when we plot them on a linear scale:



At the far left, there's a small white space below the audio spectrum representing frequencies between 0Hz and 20Hz.

Bass occupies a very small space, midrange a small space, and treble most of the space.

But when we hear sounds, our ears tell us this isn't right. Bass and midrange together account for most of what we hear, and treble is just a frosting on top. So let's replot the audio spectrum on a log scale:



Now we have something that correspond to what our ears tell us. So, when choosing how to graph something, the basic rule is: choose the method that displays the information in the most useful form, the form that best corresponds to what the information means in the real world.

End of EM2 Chapter 2

EM2 Chapter 3
AC WAVES AND WAVEFORMS

Objectives of this chapter:

- *describe common AC waveforms: square, rectangle, triangle, sawtooth, sine*
- *define and describe waveform characteristics: frequency, period, spectrum, band, phase, degrees*
- *describe "ramp" and "duty cycle"*
- *discuss fundamental and harmonic frequencies*
- *discuss the sine wave's special place in electronics, how to measure its peak and RMS voltage, and the ratios between these measurements*
- *define and discuss "noise" and "transients"*
- *discuss real-world signal waveforms and how they relate to the fundamental waveforms above*

Time to pick up where Chapter 1 left off and look at some actual AC. Some voltage patterns repeat endlessly – there are buildings where the electrical wiring has carried the same AC pattern for a century. Other signals, like audio and video, are constantly morphing. Some are random by definition, which means their signal patterns never repeat.

AC signals usually combine aspects of predictability and unpredictability, established characteristics and controlled deviations from those characteristics, and aspects that shift from one moment to the next. If we were to consider all those factors together, things would become hopelessly confused.

However, a complex signal is just a mixture of simple signals. If we study simple signals first, then when you work with a complex signal, you'll be able to analyze its components.

We're going to start by somewhat arbitrarily dividing AC signals into two groups: signals whose alternations repeat at regular intervals, and signals whose alternations do not. For simplicity, I'm only going to talk about voltages, but of course everything I say applies to current and power as well – after all, they do call it alternating *current*.

MEASURING AC

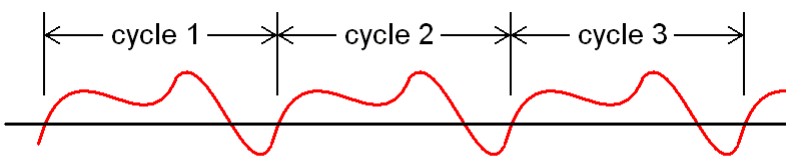
In EM1, when we were dealing with DC, measuring the voltage was easy, because it didn't change. With AC, the voltage alternates between positive and negative, which means the *average* value of an AC voltage is zero. So how do we measure it? We measure the **absolute voltage**.

You remember absolute numbers from EM1, chapter 3: a number with a line on either side of it like |50| means a number without regard to sign. |50| might mean +50 or -50. In the case of alternating current, “50 volts” means a voltage that alternates between +50 and -50. When we're talking about an AC voltage, we don't bother with the vertical lines, because we don't need to be told we're dealing with an absolute voltage; that's part of what is meant by “alternating current.” So, when you look at an appliance label that says “117VAC” you know it's talking about a voltage that switches between +117V and -117V, many times a second.

When an AC voltage represents a signal, the word that describes its strength is **amplitude**. That's kind of a mouthful, so techs more commonly use the word **level**.

REPEATING AC PATTERNS

When something repeats at regular intervals, it is **periodic**. Seasons are periodic. Monthly magazines are “periodicals.” When something is periodic we say it **cycles**, as the wheels of a bicycle do; each complete revolution is one cycle.



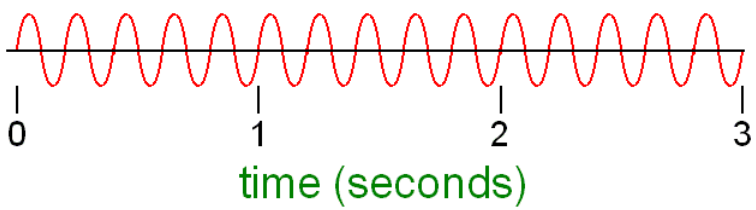
complex repeating waveform

In electronics, every time a periodic voltage pattern repeats, that's one cycle. A cycle may also be called a **wave**, especially for a signal sent through air or space. The shape of the wave is called its **waveform**.

FREQUENCY

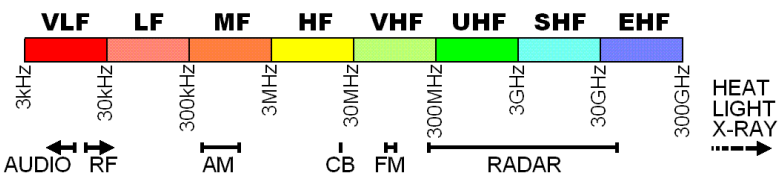
The most fundamental question we have to ask about an AC voltage, the first question that we didn't have to deal with when we were working with DC, is “How often do the alternations repeat?” How many times per second does the voltage pattern go through a complete cycle?

The number of cycles that occur in a second is called the **frequency**. Frequency used to be measured in “cps” or “cycles per second.” In 1960, the name **Hertz**, abbreviated **Hz**, was assigned to this quantity, after Heinrich Hertz who first demonstrated radio waves. The illustration below shows a waveform with a frequency of 5Hz – count the cycles.



The entire range of frequencies is called the frequency **spectrum**. The spectrum is divided into sections called **bands**, as in the AM radio band, the FM radio band, the audio band.

The lowest frequencies we deal with on a regular basis are low audio frequencies, which go as low as 16Hz. AC power in the U.S. is 60Hz, which is still a low frequency – approximately the lowest B-flat on a bass guitar. Higher frequencies use the usual prefixes – kHz, or thousands of cycles per second; MHz, or millions of cycles per second; GHz, or billions of cycles per second. The highest frequencies in use are **microwaves**, 10^9 to 10^{12} Hz. Light and X-rays are higher still. Note that the chart below shows a very wide range of frequencies, so it uses a proportional (logarithmic) scale as you'd expect.



Names have been given to various frequency bands. They are generally divided into **audio frequencies**, abbreviated **AF**, and **radio frequencies**, abbreviated **RF**. These bands are then further subdivided.

Audio-frequency signals are signals in the frequency range that human ears can hear, and a bit above and below:

- **Infrasonic** – too low to hear, below 16Hz.
- **Sonic** – described by the expression **audio band,** the range of frequencies that are audible to humans, starting at 16Hz. The top end of the audio band is usually considered to be 20,000Hz, which only very young people can hear. That's also the limit of CD audio. For most purposes, the audio band is considered to be 20Hz-20kHz. This band is further subdivided into subjective **bass**, **midrange** and **treble**.
- **Ultrasonic** – air vibrations above 20kHz. "Supersonic" is the wrong word; it refers not to frequency but to speed. Supersonic is any speed over 1,100 feet per second, the speed of sound. Ultrasonic is any frequency over 20kHz.

We're talking about air vibrations, but electrical signals in the same frequency ranges are also referred to as AF. For example, the rate at which a standard-definition television screen is scanned is just under 16kHz, which is an audio frequency.

Radio frequency signals are electrical signals whose frequencies are well above the range of human hearing. Strictly speaking, there are such things as 3Hz radio signals, which are infrasonic. But most of the time we think of RF as beginning around 50kHz.

As for the other end of the RF spectrum, the definition of the high end of the spectrum is constantly changing, as technology advances and it becomes possible to generate higher frequencies.

The RF spectrum is divided into frequency bands based on 3's. Each band has a name, shown in this chart. The chart, with additions, is duplicated in the review pages at the end of this study guide; it would be a good idea to print it out.

abbrev	*name*	*range*
VLF	**very low frequency**	**3kHz - 30kHz**
LF	**low frequency**	**30kHz - 300kHz**
MF	**medium frequency**	**300kHz - 3MHz**
HF	**high frequency**	**3MHz - 30MHz**
VHF	**very high frequency**	**30MHz - 300MHz**
UHF	**ultra high frequency**	**300MHz - 3GHz**
SHF	**super high frequency**	**3GHz - 30GHz**
EHF	**extremely high frequency**	**30GHz - 300GHz**

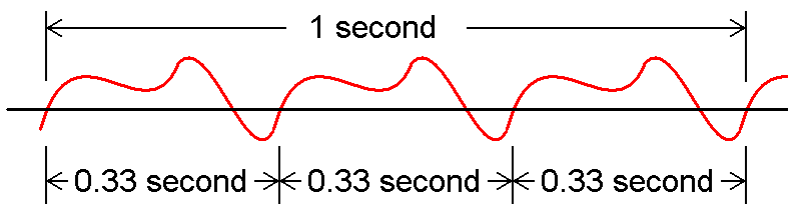
More than half of the VLF band is within the range of human hearing. It's used for, among other things, submarine navigation and communication. Only the lowest RF frequencies penetrate underwater.

At the high-frequency end, one gets the feeling they were struggling to come up with ever more flamboyant names. When the lower bands were established, nobody dreamed we'd ever have the technology to generate 300GHz.

PERIOD

If an AC waveform repeats periodically, a certain number of times a second, then obviously each cycle occupies a certain fraction of a second. This is called its **period**.

The larger the frequency, the smaller the period. In other words, frequency and period are reciprocals; for a frequency of “X” the period is 1/X. For a frequency of 3Hz, or 3 cycles per second, the period of each cycle is $^1/_3$ second.



The reciprocal works the other way too: if the period is “X” the frequency is 1/X. For a period of 10^{-5} second, the frequency is 10^5 Hz. Period is an important number when we deal with electronic control systems and data transmission.

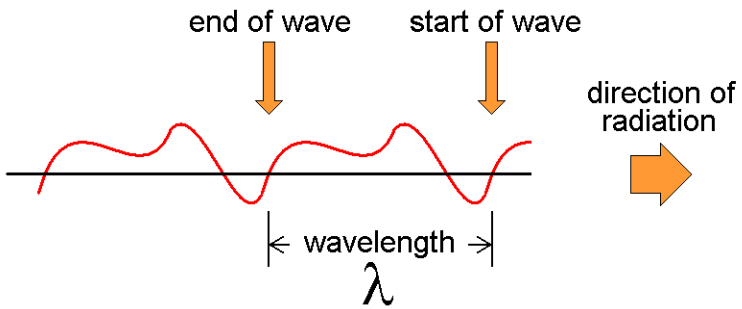
VELOCITY AND WAVELENGTH

Scientists like to use the word “velocity” when you or I would just say “speed.” The speed of an electrical signal traveling through space is the same as the speed of light, 186,000 miles per second; it's just a tiny bit slower when it's moving through air. We say the signal **radiates** through space. That's where the word “radio” came from.

In a wire or cable, we say the signal **propagates** down the length of a cable. In a cable, the signal slows down, sometimes as much as 20%. That's still darn fast, about six times around the Earth in a second. (It would be 7.4 times around in space.) You can find a cable's speed of propagation from the cable manufacturer's website.

Unless there's a need to be more precise, we just treat electrical signals as if they always travel at the speed of light. We tend to measure speed in metric units because the metric system is decimally based, as opposed to miles/yards/feet/inches. That makes it a natural fit to scientific notation, which we use a lot when talking about frequency, period and wavelength. Also, Mother Nature has given us a freebie: the speed of light and radiation just accidentally happens to be very close to a convenient **3×10^8 meters per second**. Memorize that number.

If AC waves repeat a certain number of times per second, and travel a certain distance per second, then it's easy to see that the beginning and the end of a wave will be a specific distance apart, governed by the speed of radiation and the frequency. The distance, in space, between the start of a wave and the end (which is also the the start of the next wave) is the **wavelength**. It shows up all the time in calculations involving electrical signals. It is represented by the lowercase Greek letter Lambda, λ.



Another good reason for saying “speed” instead of “velocity,” besides an admirable laziness, is that we use “V” for voltage, so we pretty much have to choose “S” for speed. Mathematically, wavelength is speed divided by frequency:

$$\lambda = S/f$$

If you look at that formula, the relationship will be obvious. The greater the speed, the greater the wavelength, because the start of one wave will be able to move a greater distance before the next wave comes along.

On the other hand, the greater the frequency, the smaller (shorter) the wavelength, because the start of one wave will have less time to travel before the next wave starts.

For example, the wavelength at 1MHz is

$$3 \times 10^8 / 10^6 = 3 \times 10^2$$

That is, a speed of 3×10^8 meters per second divided by a frequency of 10^6 Hz works out to a wavelength of 3×10^2 meters, or 300 meters, roughly 1,000 feet. One practical use for this formula is calculating the proper length of an antenna, which is related to the wavelength of the signal it carries. Did you know that the height of an AM radio station's antenna depends on the station's frequency? You do now.

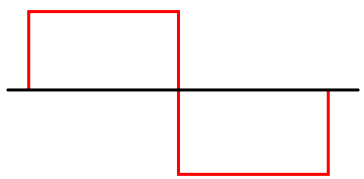
One of the reasons the various bands' frequencies are marked off by 3's is that it conveniently divides up the wavelengths in multiples of 10. Radio techs talk about the "10 meter band," the "100 meter band" and so on. These concepts are worth a tech's time to learn. It will be handy to be able to look up in the sky and visualize the wavelengths of the signals you're dealing with.

Suggestion: Get a meter stick and hang it up on your wall. Cut pieces of paper to lengths of 10cm, 1cm and 1mm and hang them up next to the meter stick. Look at them a lot. You need to develop a good feel for just how long these dimensions are, cuz you'll be using them. While you're at it, hang up yard, foot and inch sticks too. You will need to get a feel for mental foot/meter conversions.

WAVEFORMS

We know AC consists of alternating positive and negative voltages. That doesn't describe *how* they alternate. They may change smoothly or abruptly, or spend more time in one direction than the other. Any wave shape or waveform can be generated to fit a particular application.

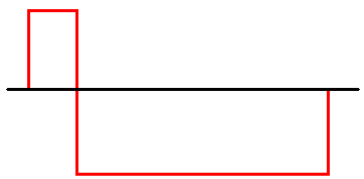
Common waveforms have names that describe their shape. The most common are the square wave and its near cousin, the rectangular wave; the trapezoidal wave, the triangle wave and its cousin, the sawtooth wave; and the sine wave.



The simplest waveform, electrically, is the **square wave**, which is called that for obvious reasons. The voltage waveform is positive for a certain time period, then abruptly goes negative for an equal time period, then goes positive again. The associated current flows one way for a while, then abruptly reverses and flows the other way before reversing again.

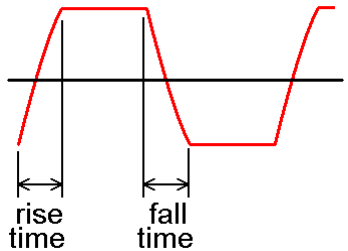
In a square wave, the positive and negative voltages are equal. For example, if the positive part of the wave is +3 volts, the negative part will be -3 volts, and we'd call this a 3V square wave.

In a square wave, the positive/negative time ratio is always 50:50. Another way of saying this is that a square wave has a 50% **duty cycle**.

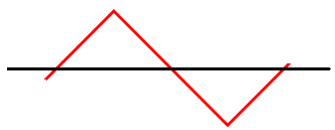


If the positive and negative durations are not equal, the square wave becomes a **rectangular wave**. For example, if you see a blinking light, and it's off a lot more than it's on, you could say it is being driven with a very low-frequency rectangular wave. For a rectangular wave, the duty cycle can be anything that is higher than 0% and lower than 100%.

In a perfect square or rectangular wave, the transitions between positive and negative would be instantaneous, accounting for 0% of the period of the wave. In practice, the voltage takes a small amount of time to go positive or negative. These times are called **rise time** and **fall time** respectively and they are hugely important with high speed digital communications, as you'll see in EM4. For high digital data rates, the rise and fall times must be as short as possible. On the other hand, where high data rates are not involved, rise and fall times are often deliberately lengthened to reduce the amount of high-frequency interference they can radiate.



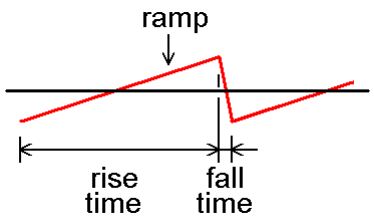
If rise and fall times are extended to the point where they take up a considerable part of the wave, the wave is called a **trapezoid wave**. Some DC-to-AC inverters use it to crudely imitate a sine wave. For many AC power applications, a crude imitation is good enough.

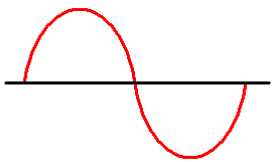


A **triangle wave** consists of a period of smoothly rising voltage followed by an equal period of smoothly falling voltage. When voltage rises or falls at a uniform rate, that part of the waveform is called a **ramp**.

If the duty cycle is not exactly 50%, it's still considered a triangle wave, but if it's way different – e.g. if there's a comparatively long period of slowly rising voltage, followed by a comparatively short period of quickly falling voltage – it becomes a **sawtooth wave**. The theoretical limit is a 100% duty cycle, where the rise time takes up 100% of the cycle and the fall time takes 0%.

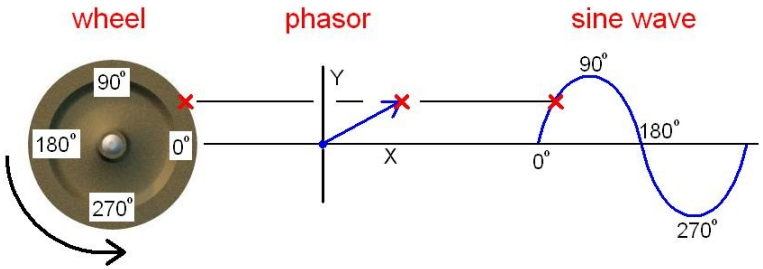
That is an idealized waveform that can't be achieved in practice, but 99.99% / 0.001% is not much of a challenge. A sawtooth can also take the opposite form, with a long fall time and short rise time.





A **sine wave** is a smoothly changing shape with broad tops and bottoms which some people call **peaks** and **valleys** and others just call positive and negative peaks. You've seen pictures of sine waves before, but have you ever thought of where they come from or why people use sine waves? The answers to those questions are important. We're going to spend some time with this, because the sine wave is *the* fundamental AC waveform. It's the basis of all other waves and by far the most often used for testing in analog electronics.

The sine wave displays what is called **pure harmonic motion**, graphed over time. If you take a wheel with a perfect, friction-free bearing, put a mark on the rim, and spin it, the mark will move in two dimensions – up-and-down and side-to-side. If we only look at the up-and-down motion, and graph that motion over time, we get a sine wave. The name refers to the mathematical function that describes it, which includes a sine term.



The mathematical equivalent of a wheel with a mark on the rim is a phasor that rotates around the origin of an X/Y graph. It is customary to start with the phasor pointing to the right, along the X-axis. That way the length of the phasor is a simple positive real number. The phasor rotates counter-clockwise through 360 degrees per cycle. Suppose the length of the phasor is the same as the voltage of a sine wave at its highest point. That means, when the phasor is 90 degrees from where it started – when it's pointing straight up – its point is as high on the Y-axis as the peak voltage. And at every other position, its Y-axis height is the same as the sine wave voltage at that position. In other words, the Y-coordinate of the phasor point represents the instantaneous voltage at *every* moment.

This gives us a very useful tool. You recall from the last chapter's discussion of trigonometry that the three sides of a right triangle are related by ratios called “sine,” “cosine” and “tangent.”

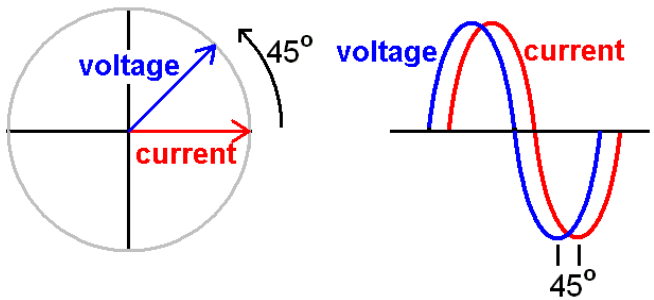
We already know that one of the angles in a right triangle is 90 degrees. If we know one of the other angles, and the length of one of the sides, we can use the known sine, cosine and tangent ratios to find the other two.

Here's how it applies to sine waves. As the phasor rotates, it forms an angle with the X-axis that is constantly changing. We know the phasor sweeps through a certain fraction of the 360-degree circle in a certain time. The phasor's rotation is smooth – remember that perfectly friction-free bearing – which is to say that it always rotates through exactly the same number of degrees per second, with no bumps or glitches. If it takes one second for the phasor to rotate the full 360 degrees, corresponding to a frequency of 1Hz, then we know that in $^{1}/_{10}$ second it rotates through 36 degrees. Thus we know the phasor's angle at every moment. If we also know its length, we can use trigonometric ratios to find the phasor tip's Y-coordinate at any instant. If the phasor's length represents the peak voltage, its Y-coordinate at any instant represents the voltage at that instant.

Because the sine wave is related to circular motion, we speak of a position at some point during a cycle in terms of **degrees**. The point where the sine wave crosses the X-axis going in the positive direction, at the instant when the voltage is zero, is zero degrees. The positive peak is 90 degrees, the negative-going zero crossing is 180 degrees, and the negative peak, or valley, is 270 degrees.

PHASE

One thing that happens a lot in electronics is that we deal with two or more phasors at the same time. In a given circuit, for example, current may not be precisely in step with voltage – current peaks may occur before or after the voltage peaks. Phasors give us a good way to show what's going on.



The word **phase** usually refers to the relative positions of two waveforms, stated in terms of degrees. For example, suppose we're dealing with a 1Hz sine wave and, at a certain point in a circuit, the current waveform lags behind the voltage waveform by 0.125 second. At 1Hz, the period representing a full 360-degree cycle is 1 second, so a 0.125-second lag would represent 360 x 0.125 = 45 degrees. We would say the current **lags** the voltage by 45 degrees; if it were ahead of the voltage, we would say it **leads**.

It's important to remember that a phasor diagram represents a snapshot of a *moving* phasor at a particular instant.

We'll talk more about phase later. Right now, we're going to talk about why the sine wave is so important in electronics. We'll start with a digression into music. Bear with me, this sine wave discussion does come full circle (pun intended).

SINE WAVES AND HARMONICS

Every musician knows that a tone is usually made up of a combination of tones. There is the **fundamental**, which is the basic pitch of a note, like middle C. And there are **overtones**, which occur at higher pitches that are related to the fundamental – an octave higher, an octave and a half higher, and so on.

We don't normally perceive these overtones separately, they are blended into the overall tone. But they are still important. A trumpet playing a certain note sounds different from a clarinet playing the same note because the amounts and distribution of the overtones are different. In other words, overtones are responsible for musical tone colors. A highly trained listener can tease a complex tone apart and hear its components individually.

What a musician hears as "pitch" is related to frequency. Frequencies that are perfect multiples of the fundamental pitch, which musicians call overtones, are called **harmonics** by technical people – and yes, the term is related to musical harmony. Strictly speaking, all the related frequencies in a complex waveform are harmonics, including the fundamental. However, when we talk about harmonics, we usually mean everything *but* the fundamental. It's incorrect, but it does give us a way to describe the difference between the fundamental frequency of a complex waveform and all the other frequencies that are part of it.

In a regular, unchanging waveform, the harmonics are exact multiples of the fundamental. If the fundamental is 100Hz, the harmonics will fall on 200Hz, 300Hz, 400Hz, 500Hz and so on.

5th harmonic	=	500Hz
4th harmonic	=	400Hz
3rd harmonic	=	300Hz
2nd harmonic	=	200Hz
1st harmonic fundamental	=	100Hz

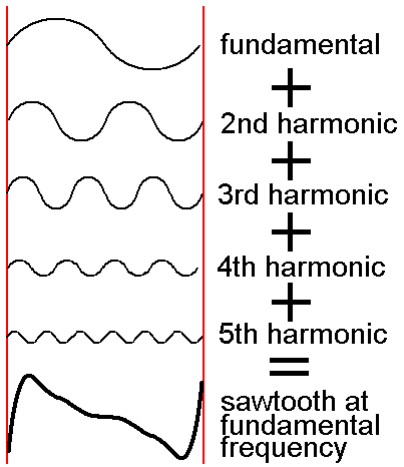
Harmonics are numbered, and each number corresponds to the particular multiple of the fundamental frequency. For example, low "G" on a lead or rhythm guitar – the fundamental pitch – is around 100Hz, which strictly speaking is the 1st harmonic. The 2nd harmonic, then, is twice the fundamental or 200Hz, the 3rd harmonic is three times the fundamental or 300Hz and so on.

Lower harmonics, like 3x the fundamental, are called **low-order** harmonics. Higher ones, like 30x the fundamental, are called **higher-order** harmonics.

I began by speaking of musical tones, which implies that I'm talking about audio frequencies. But the same principle applies to waveforms at any frequency. In fact, one of the ways of generating a high radio frequency is to generate a lower frequency and then use a frequency-selective circuit to pick out and reinforce one of its harmonics.

Now back to sine waves. The sine wave is important in science and electronics because *it is the only waveform that has no upper harmonics*. It's pure, 100% fundamental. Every other waveform is a mix of fundamental and harmonics – in other words, a mix of sine waves at different frequencies. A 100Hz sawtooth wave has a bit of 200Hz in it, a bit of 300Hz, 400Hz, 500Hz and so on. As a French scientist named Fourier discovered, every repeating waveform, no matter how complex, can be sorted out into sine waves of differing strengths at the fundamental and upper harmonic frequencies.

This is shown in the illustration above right. We start with a pure sine wave at the fundamental frequency, then add a little sine wave at 2 times that frequency, then 3, 4 and 5 times the frequency. Note that the composite waveform at the bottom resembles a sawtooth wave.



If we kept adding harmonics in the right proportions up through infinity, we'd wind up with a perfect sawtooth wave with zero rise time, built from nothing but sine waves.

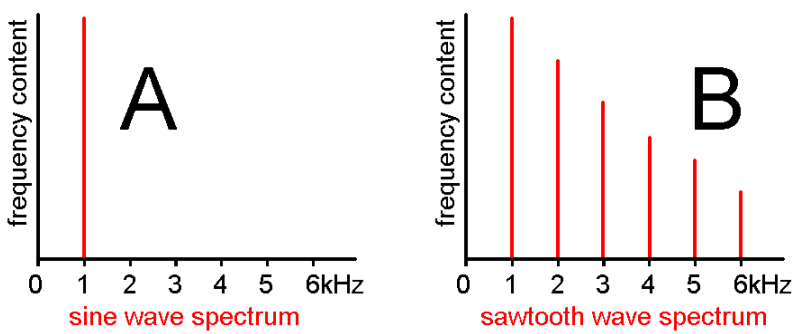
This brings us to an important point: the faster the rise and fall times of a signal, the more high-order harmonics and high-frequency energy are present. When we build a sawtooth wave, or for that matter a square or rectangular or any other kind of wave with steep slopes, the fast rise and fall times mean there's a lot of high-order harmonic energy in the wave.

Here are two real-world implications. To reduce high-frequency radiation, light dimmers have to include components to stretch out rise and fall times or they'll radiate enough RF to mess up radio and TV reception. On the other hand, long cables reduce high frequency energy, slowing down rise and fall times. This makes it difficult to transmit high-speed digital data over long cables. At the moment this is causing a lot of problems with HDMI video, a technology that pushes cables to their limits.

As noted earlier, the word “spectrum” refers to the whole range of frequencies, from audio to light and beyond. There's another definition for that word. The whole range of frequencies contained *in a waveform* is called the **spectrum of the wave**.

SPECTRUM ANALYSIS

A useful way to show the harmonic spectrum of a waveform is with a special kind of graph where the X-axis, instead of corresponding to time, corresponds to frequency. Lower frequencies are to the left, often but not always with zero Hz at the origin; higher frequencies are to the right. Figure A shows the spectrum of a 1kHz sine wave, which doesn't contain energy at any frequency except 1kHz. Figure B shows the low-order harmonics of a 1kHz sawtooth wave; observe that it contains energy at every multiple of 1kHz.



SPECTRAL PURITY

In many cases, it is important for an AC voltage or current of a certain frequency to be that frequency alone, and nothing else. For example, if an AM radio station is transmitting at 630KHz, and the signal is not a pure sine wave, it will also be radiating energy at 1260KHz, which may interfere with another station that actually belongs on that frequency. For this reason, a radio station's signal must meet legally established benchmarks for harmonic purity. In other words, it must be a true sine wave.

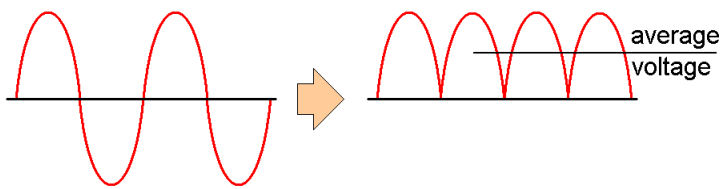
Sine waves are routinely used for testing. If you wanted to test an amplifier's behavior at 1kHz, why would you feed it a test signal that is a mix of 1kHz, 2kHz, 3kHz and so on? How would you know what the results meant? That's why sine-wave testing is so common – not because pure, continuous, unchanging sine waves closely resemble real-world signals (they don't) but because, if the test signal isn't a pure sine, you'll have trouble figuring out what your test results mean.

The sine wave we deal with most often is the AC power a building is wired with. It leaves the power plant as a fairly pure 60Hz sine wave, though it gets screwed up a bit by the time it reaches the end user. Lamp dimmers in particular are famous for putting glitches on the AC sine wave.

MEASURING THE VOLTAGE OF A WAVE

Time to talk some more about measurement. Consider the sine wave. How would you measure its voltage? You could measure it at the peaks, but that's not very realistic, because the sine wave spends very little time there. For most of the sine wave, the voltage is considerably less than its peak voltage, and it's zero twice in every cycle.

It would make more sense to measure the average voltage. In concept, that's easy. You flip over the negative alternations, so the voltage is always positive; this takes care of that "absolute value" stuff.



Then you look at where the wave spends most of its time and draw a line. There's a simple electrical circuit, which we'll study in EM3, that can be used to obtain this average.

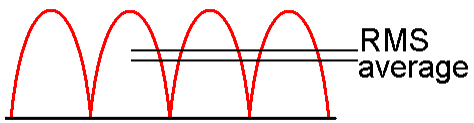
The problem is, when we start using this average to make power calculations, we get an error. This is due to the fact that the power at any instant is related not to the voltage but to the *square* of the voltage. And if the waveform is something weird like a 10%-duty cycle rectangular wave, the power calculation will be nowhere close to right.

RMS VOLTAGE

Since we're dealing with a square relationship, what we have to do to bring voltage and power calculations into harmony is to measure the voltage at many points along the wave, then separately square each of these voltages and calculate the average or "mean" of all those squares. Finally, we take that average and find its square root. This gives us the true voltage. Doing this mathematically is tedious, but there are electrical circuits that can square and square-root a voltage in real time. The result is called the **Root-Mean-Square** or **RMS** voltage, the root of the mean of the squares. RMS voltage is also called the **effective** voltage.

By the way, when you see amplifier specifications like "100 Watts RMS," that's technically wrong. Voltages and currents are squared, not power. But everybody talks this way, and if we want to cut them a little slack, we could say "100 Watts RMS" means a wattage that's calculated from an RMS voltage measurement and a known load resistance. Power is often calculated that way, instead of using a wattmeter. Anyway, there are more important things to complain about, like the fact that every new generation of cellphones sounds worse than the previous generation, despite advances in technology. Guess who's been phone shopping recently.

It's worth noting that most AC voltmeters do not measure RMS voltage, which is a bit complicated to measure, but average voltage, which is easy. With a sine wave, the difference between the average and RMS voltage is about 10%, so the meter manufacturer just measures the average voltage and then fudges the result upward about 10% to get a reading that corresponds to the RMS voltage. With a sine wave, this works and is perfectly legitimate.



However, when the waveform isn't a sine wave, the reading will be inaccurate. To measure a voltage that isn't a sine wave, you need a true RMS-measuring (not just RMS-reading) meter. This kind of meter costs more because of that squaring and square-rooting circuit, which is why only deluxe meters have it.

Moreover, most meters don't flip over the negative alternations, but just discard them, which causes even more trouble when you're trying to measure a wave where the positive and negative alternations aren't alike. And to add insult to injury, most meters don't have linear frequency response. Lower-priced DVM's, especially, tend to droop above 200Hz.

You can measure any waveform you like with any old meter, but you should regard the result as just a ballpark estimate unless you're measuring a 60Hz sine wave, where you know the reading is accurate, or you're using a deluxe meter with wide frequency response and RMS circuitry.

When we talk about a sine wave's RMS voltage, we usually don't bother saying “RMS.” In the case of a sine wave, RMS voltage is just assumed unless we specify that the number refers to some other kind of measurement.

PEAK VOLTAGE

I don't have nearly as much to say about measuring the **peak voltage**, because it's simple. You just measure the highest instantaneous voltage peak of the waveform.

I should say, it's simple in concept. In practice it's a little harder, because meters – especially mechanical meters with d'Arsonval movements – respond to average voltage. A peak-reading meter requires some circuitry, along with a battery to power it. If you need to know the peak voltage of a waveform, it's often easier to measure the RMS voltage and then apply a known formula to find the peak.

This depends on the waveform, and it takes a really clever tech to analyze some of the more esoteric ones. But the peak-to-RMS ratios for sine waves are well known, and you should memorize them.

PEAK-TO-RMS CONVERSION OF SINE WAVES

In the EM1 Math chapter, I said you'd often be using the square root of 2, which is about 1.4, and the square root of ½, which is about 0.707. This is one of those cases.

In a sine wave, the peak voltage is the square root of 2 times the RMS voltage. For example, a 10-volt sine wave has a peak voltage of about 14 volts. Going the other way, the RMS voltage is the square root of ½ multiplied by the peak voltage. A sine wave that peaks at 10 volts is a 7.07-volt sine wave. If we were calculating current, these same ratios would apply.

For instantaneous power, the peak-to-RMS ratios are the voltage ratios squared. (Big surprise, right?) If you take the square root of 2 and square it, you will of course get 2. Likewise, if you square the square root of ½ you'll get ½. So the instantaneous power of a sine wave when the voltage is at its peak is twice the average power, and the average power is half the peak power. Memorize the following:

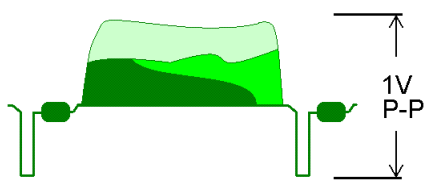
PEAK VOLTAGE = 1.4X RMS VOLTAGE
RMS VOLTAGE = 0.7X PEAK VOLTAGE
PEAK POWER = 2X AVERAGE POWER
AVERAGE POWER = 0.5X PEAK POWER

In the case of power it is correct to speak of "average" rather than "RMS."

This has an important safety ramification: where dangerous sinewave voltages are concerned, keep in mind that the peak voltage is higher than the RMS. For example, I said on page two that 117VAC power wiring alternates between +117V and -117V. That is true from an RMS standpoint, but the peaks reach almost 170V. You should assume a high AC voltage is even more dangerous than it looks.

PEAK-TO-PEAK VOLTAGE

Peak and RMS voltage measurements are both useful. We also sometimes use a **peak-to-peak** or **P-P** voltage measurement to describe the entire voltage swing from maximum positive to maximum negative voltage. This is important when we want to specify the voltage range an amplifier must handle. It's also useful when the waveform is complex. A video waveform, for example, spends most of its time going up and down within certain limits, in a fairly unpredictable fashion, with periodic negative-going spikes outside those limits.



It makes no sense to talk about the RMS value of this waveform, or of the peak voltage – peak in which direction? The only thing that matters is that the circuitry can accommodate the whole waveform, from its most positive point to its most negative. So a typical video waveform is described as having about a 1-volt amplitude, peak to peak.

“Peak-to-peak power” is a nonsense expression, a contradiction in terms. Only marketing hypesters use it, to man-up wimpy amplifiers in advertising literature. I hope said hypesters are properly ashamed.

There is one AC waveform where the peak, average and RMS values are identical: the square wave. Since the absolute voltage does not vary – it's always the same voltage, either positive or negative – it can be thought of as a DC voltage that reverses polarity. Note that a standard multimeter will read about 10% high on a square wave, since it's built with the average-to-RMS fudge factor for sine waves.

NON-REPEATING WAVEFORMS

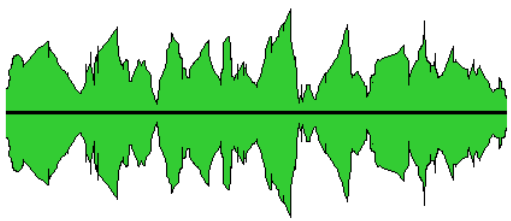
Non-repeating waveforms are useful, but they can also be a pain. Nature generates its own voltages, which take the form of non-repeating waveforms. They are random, and when they get into our circuitry, they cause problems.

NOISE

In electronics, “noise” has two definitions. The definition we're interested in at the moment: noise is a voltage

- whose instantaneous value is random, and
- therefore cannot be predicted, and which
- contains energy at every frequency across the entire spectrum.

(The other definition: noise is anything that contaminates or degrades a signal. We'll deal with that in chapter 10.) If you tune an old analog TV to a nonexistent station, the rushing sound you hear and the snowy screen you see are caused by random electrical noise.



In the illustration above, the variations you can see are low-frequency noise. The green area represents mid-and high-frequency noise, voltage fluctuations so fast that they visually blend together. That's what a noise signal containing energy at every frequency looks like.

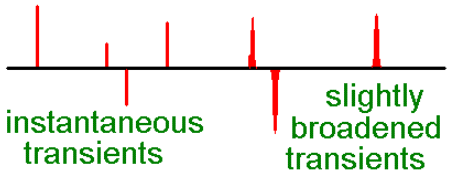
We can't talk about the frequency spectrum of noise in terms of “cycles” because noise doesn't cycle; there is nothing about any aspect of noise that repeats from one moment to the next. However, though the noise voltage is random in the short term, we can usually determine what its long-term *average* will be with a high degree of certainty. This allows noise to be used as a reliable test signal. Because it's constantly changing, and has energy across the spectrum, it can often explore the behavior of a piece of equipment more thoroughly and more quickly than sine, square or triangle wave testing can.

To measure noise, a true RMS-measuring meter is absolutely necessary.

TRANSIENTS

A **transient** is a single, short-duration change away from zero volts and back again. In theory, a transient has zero period (it is instantaneous) and contains energy all across the frequency spectrum, as well as DC. In practice, the definition of a transient is broadened a bit to include any electrical event that:

- consists of a single short-duration voltage spike (or, if it repeats, the repetitions are far apart), and
- contains energy over a broad enough frequency range that no single characteristic frequency can be determined.



The reason for fudging the definition is that, if we insisted on the precise definition, no real-world transient would fit it. No voltage change can have infinitely fast rise and fall times or zero duration or contain energy at infinitely high frequencies. So, along with “snap,” we allow “whack” and “thump” to be called transients. The sound of a kick drum counts as a transient: it is of short duration, surrounded by silence (of the drum, not the band) of much longer duration, and depending on how the drummer has damped it, you can't pick out a particular note or pitch that it sounds. In the context of continuous, pitch-related sound generators like guitars and keyboards, the sound of a kick drum counts as a transient, even though it contains little energy above 200Hz. On the other hand, the spark of a furnace igniter contains energy up to hundreds of MHz.

With a transient, it is probably useless to speak of the RMS voltage; the peak voltage is what matters.

Speech, music, video and many other kinds of signal contain transients.

Noise and transients are related. Noise is just transients of random intensity, occurring so frequently that they blend together into a continuous, randomly varying AC voltage. The sound of a raindrop hitting a tin roof is a transient. The sound of a million raindrops hitting a tin roof is noise.

PROBLEMS CAUSED BY NOISE AND TRANSIENTS

When Nature gifts us with a transient, analog signals may be disrupted, digital data may be corrupted, and in extreme cases, equipment may be damaged.

With very weak signals, one of the things that gets in our way is the fact that some noise is *always* present; there is actually no such thing as zero AC volts in a circuit. The noise comes from the random motion of charged subatomic particles and the electromagnetic energy they generate. This constant, low-level AC voltage is called the **noise floor**. There is always a limit to how weak a signal can get before it's degraded by noise or lost below the noise floor entirely, and once noise becomes part of a signal, it's usually impossible to remove it.

A considerable portion of the effort involved in electronic technology goes to keeping transients and noise away from signals. As proof of this, we will be dealing with it in chapters 4, 8 and 10, and every one of the remaining EM6 study guides.

COMPLEX WAVEFORMS

As a house is assembled from boards, shingles, tile, carpet, glass and so on, all the complex real-world signals in electronics are made of waves of various shapes and frequencies, and noise and transients.

The transmission of information involves change; thus an unchanging, completely repetitious waveform is by definition incapable of transmitting information. On the other hand a completely random, chance-driven and thus uncontrolled signal is also by definition incapable of transmitting information. An information-bearing signal has to be a mix. It must follow a pattern so the equipment that deals with it can recognize it as a signal, but the pattern must be changeable within specified limits so information can be communicated.

Many signals are pretty complex. But, as stated earlier, complex waveforms are just composites of basic waveforms. You'll meet these real-world signals in EM3 through EM6, and with the knowledge you now have you'll be able to analyze them. But, for the remainder of this study guide, we're going to continue to treat AC waveforms not as information-bearing signals but purely as waveforms. In the next chapter, I'm going to introduce you to some test gear that allows you to create, view and analyze AC waveforms in real time.

End of EM2 Chapter 3

EM2 Chapter 4
TEST EQUIPMENT AND MEASUREMENTS

Objectives of this chapter:

- *list all the controls of a standard oscilloscope and describe each of their functions*
- *describe the signal path from input to display of an oscilloscope*
- *discuss how you would set an oscilloscope's controls to display various kinds and levels of signal*
- *explain the function and use of a spectrum analyzer*
- *describe the operation and uses of the following signal generators: analog sine wave, function, look-up table, beat frequency, sweep, white noise, pink noise*

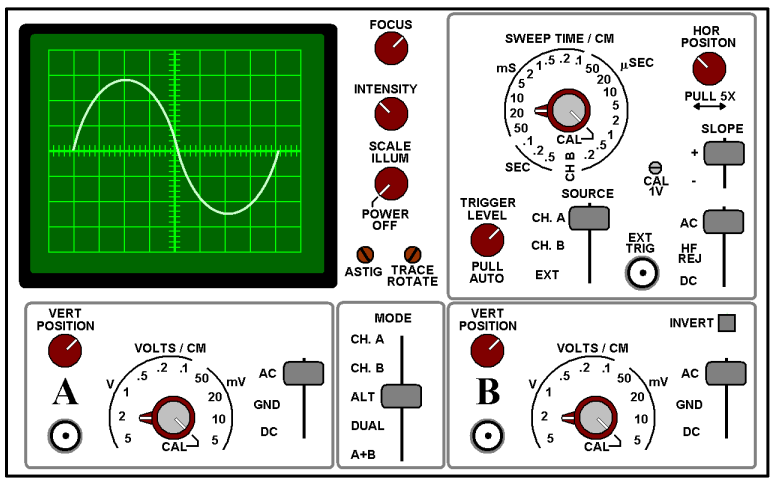
It will come as no surprise that AC test gear mostly consists of equipment to generate and display time-varying voltages. We're going to look at display equipment first, cuz ya gotta be able to see the road before you can drive on it.

THE OSCILLOSCOPE

In EM1, we spent most of the "test equipment" chapter looking at the multimeter or VOM (volt-ohmmeter), which is *the* basic piece of DC test gear.

The basic piece of AC test gear is the **oscilloscope**. It is your window on what's going on inside a circuit. A tech who hasn't learned to use it is effectively blind. As you'll see, this chapter goes into very great detail on scopes, how they operate, and how they're used. I'm going to go through all the controls on a typical scope. If you've got one, look at it while you're reading this.

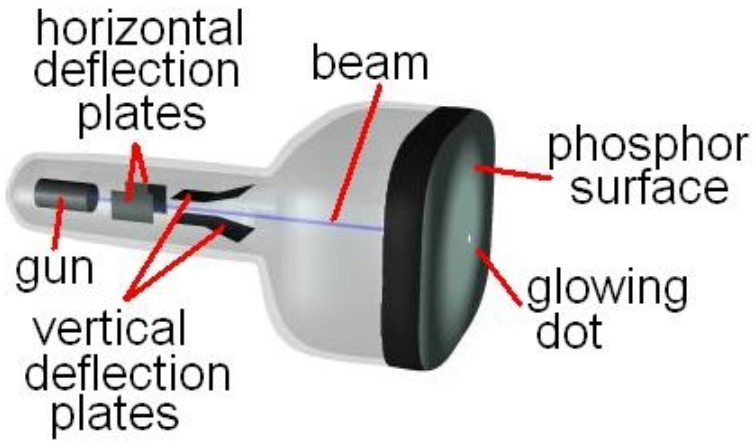
Scopes come in two flavors: analog and digital.



I'm going to use an analog scope as an example. I've gotten a lot of conflicting advice on this. Some think the analog scope is a dinosaur. Others point out that they're still being manufactured and sold, and that a top-quality used analog scope does many things better than a digital scope of the same price. My analog bench scope was bought on Ebay for $300, and it will display waveforms up to 275MHz, which a $1,000 digital scope will not do. On the other hand, digital scopes have waveform capture and analysis tools no analog scope can touch, and they're more compact.

I finally decided to tilt most of this discussion in the direction of analog scopes, for three reasons. One, they're easy to understand. Two, I'd like to think you might buy a scope of your own, and used analog scopes are cheap enough for a student to afford. Three, basic digital scope functions mimic analog scope functions, and I can return to digital scopes in EM4, after you've learned about digital technology.

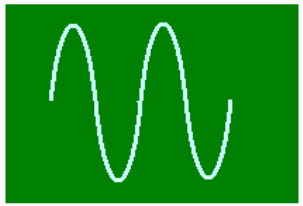
The classic analog oscilloscope is based on a **cathode-ray tube** (**CRT**) display tube with no air inside. An electron gun at the back shoots a beam of electrons at the face of the tube, which makes a phosphor coating on the inside of the glass light up. If left alone, the beam would light up (and probably burn a hole in) a single spot in the middle of the screen. However, in normal operation the horizontal deflection plates bend the beam on its way to the face so as to scan left to right, creating a horizontal line. In this way it is almost identical to a television CRT. The difference is



that, in the TV CRT, the beam repeatedly scans the whole face, drawing a 2-dimensional picture. In an oscilloscope, the beam traces a single line from left to right.

Suppose you took a pencil and started to draw a line from left to right across a paper, moving at a constant rate. Suppose, halfway across, someone bumped your hand toward or away from you, but did not bump it sideways, so as not to disturb your steady progress across the paper. When your hand got to the right side of the paper, you'd have a visual record of the event.

That's what the beam in a scope CRT does. Its left-to-right movement goes at a constant rate. Its up-and-down movement corresponds to instantaneous signal voltage at the scope's input. The result is a line, called a **trace**, that displays the signal's voltage versus time – in other words, its waveform.

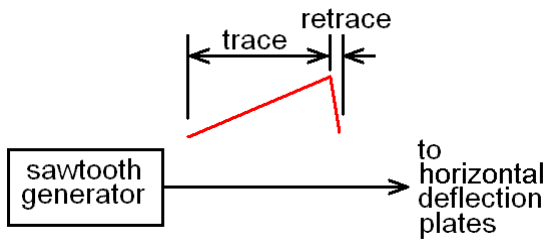


RETRACE

In chapter 3, when I showed a repeating waveform, I showed a single cycle. In a diagram, that's all that needs to be shown, since every wave is the same as every other. However, when we look at a waveform on an oscilloscope, we're not looking at a theoretical representation, but a signal occurring in real time. If it changes, we want to know. So the scope has to continually display what's going on.

Suppose we want the beam to move fast enough so a single cycle of 1kHz sine wave appears half an inch wide on the CRT face. This means the beam must move half an inch per millisecond horizontally, or 500 inches per second. If we want to watch a minute's worth of that, we need a screen that's 2,500 feet wide - upwards of half a mile.

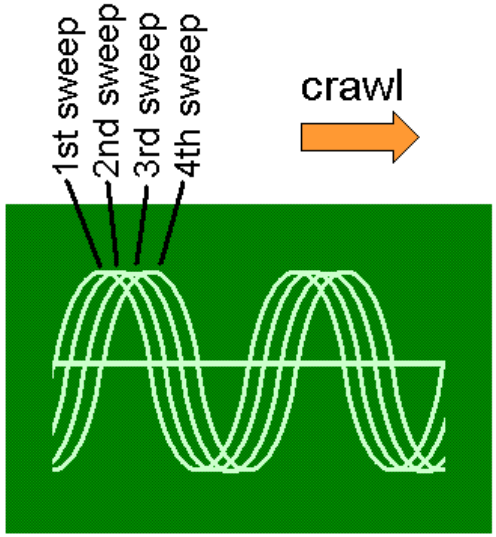
How do we fix that problem? We build the scope so every time the trace reaches the right side of the screen, it snaps back to the left and repeats the same left-to-right motion, over and over again. If the signal waveform hasn't changed, and the scope's scanning frequency is the same as the signal frequency, the beam will trace the same path again. When the signal waveform does change, we'll see it. It isn't difficult to make a scope do this. We just build a sawtooth generator and make its waveform control the horizontal position of the electron beam. During the slow upward ramp of the sawtooth, the beam sweeps from left to right. When the sawtooth reaches the fast downward ramp, the beam snaps back to the left. This snapback interval is called the **retrace**.



BLANKING

We don't want the retrace interval to show on the screen. That would put two lines on the screen, one going the wrong way, which would make for a confusing display. So we add a circuit which shuts off the electron beam during the retrace interval. This is known as **blanking**.

SYNCHRONIZATION

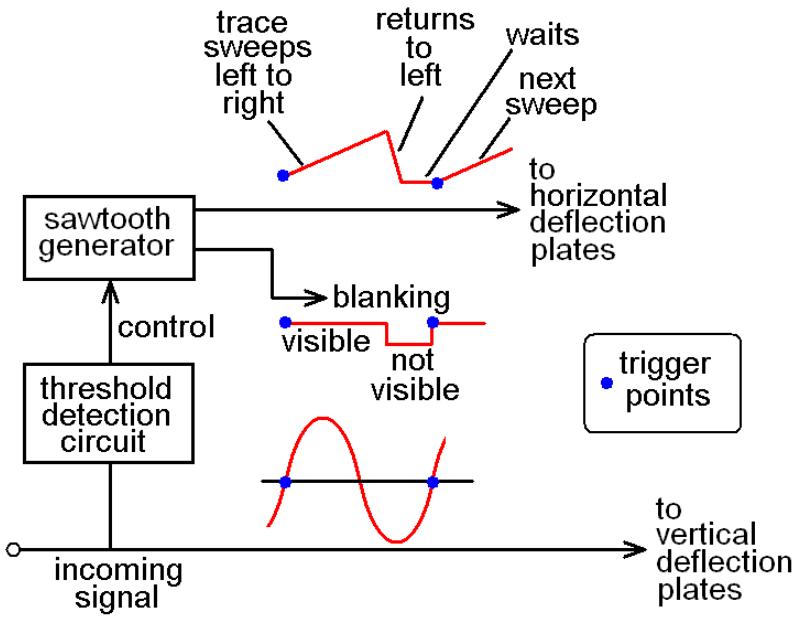


The biggest remaining problem is synchronizing the sweep frequency – the frequency of the sawtooth oscillator – with the incoming signal frequency. Suppose we want to look at a 100Hz sinewave, but the sawtooth oscillator is running at 97Hz. That means the second sweep will start late, when the sinewave is already 3% through its cycle. On the third sweep, the wave will start 6% ahead of the sweep, on the fourth sweep 9% ahead and so on. This will shift the position of the displayed wave a little with each sweep, making it crawl across the screen.

The displayed signal waveform will only stand still on the screen if the sawtooth generator is running at precisely the same frequency as the signal. At high frequencies, say 10MHz, it's almost impossible to manually adjust the sweep frequency precisely enough to get a stable display.

THE TRIGGERED SWEEP

We solve this problem by adding a threshold circuit that monitors the instantaneous voltage of the input waveform, looking for the moment when it crosses the zero-volt line, going from negative to positive (blue dots in the diagram below). When the beam completes a sweep, it snaps back to the left, but waits for the threshold detection circuit to tell it to start the next sweep.



This is called sweep **triggering**, and refers to the fact that each sweep is triggered when the signal crosses the zero line going in the positive direction. If the waveform doesn't change from one sweep to the next, the trigger will hit at the same point in each wave and the electron beam will keep tracing the same wave shape in the same position, resulting in a stable display.

This eliminates the need to synchronize the sawtooth sweep generator with the signal frequency. The electron beam traces the signal waveform for as long as it takes to cross the screen, then it returns to the left and waits for the trigger circuit to tell it to start the next sweep.

You, the user can set the sweep to whatever speed you like. This is an especially useful feature because many real-world signals change in frequency or are a complex mix of frequencies, and couldn't be synchronized any other way.

Adjustable Triggering

We may not always want triggering to occur at the point where the signal waveform crosses zero. For one thing, some complex waveforms cross zero more than once each cycle.

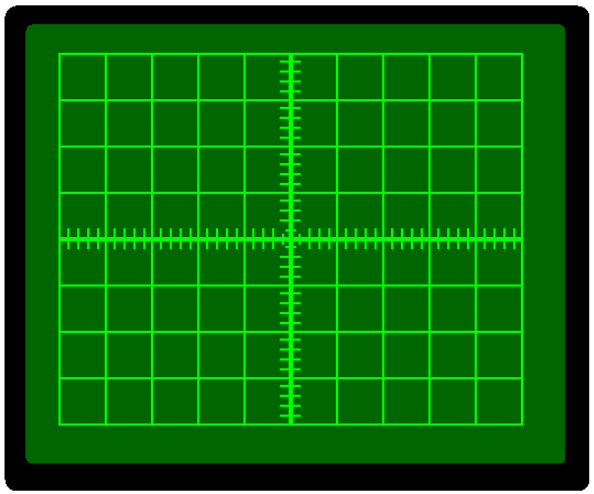
An **adjustable trigger** allows us to set the threshold voltage. Most scopes also allow us to flip the trigger polarity over, so the sweep triggers when the waveform crosses the threshold going in the negative direction.

Automatic Triggering

Most scopes offer the option of **automatic triggering**. This is a circuit that monitors the input voltage and automatically sets the threshold so that, no matter how large or small the signal, there will always be a point where the sweep is triggered. This is handy because, if the threshold does not happen to be set to zero, a small signal may not trigger a scan at all, and the screen goes dark. Not a crisis, but it is an annoyance to have the trace disappear when the signal voltage is low.

Most techs leave the triggering set on automatic most of the time, and only turn to manually-adjusted triggering when they're dealing with a funky waveform that the automatic circuit doesn't know how to handle. Or, they may choose manual triggering when they want to closely examine the waveform right where it crosses the zero line, so they need the sweep to be already underway by that point.

SCALE GRADUATIONS



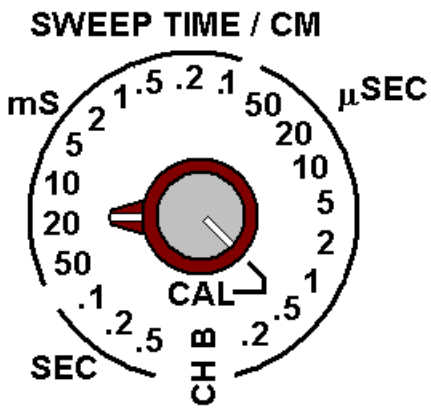
The oscilloscope has horizontal and vertical lines on the face which form a grid of squares. This illustration shows a scope with 10 major graduations in the horizontal direction and 8 vertical ones. The X- and Y-axes have 5 minor graduations per major one. The major graduations are typically one square centimeter each, a common dimension for a full-sized bench scope.

If we know how fast the trace sweeps across the screen, the lines can be used as reference points to visually determine the period of the signal waveform. A quick mental calculation if you're macho, a few pokes at the calculator if you're not, gives you the frequency. Likewise, the gradations can be used to determine the signal voltage at any point in the wave. I'll show you how in a bit.

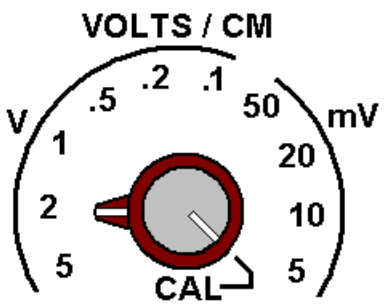
HORIZONTAL RANGE SWITCHING

A trace ought to be able to move so slowly that it can display signal voltage changes that take several seconds to happen, or so fast that it can show the highest frequencies the scope is capable of handling. In a typical scope, the slowest sweep takes five seconds to go across the screen; the fastest sweep occurs in a couple of microseconds.

Sweep speed is typically selected by a rotary switch calibrated in increments of 1, 2, 5, 10, 20, 50, 100, 200, 500 and so on. The slowest speed might be half a second per major division on the scope face. If the scope face has 10 divisions, the trace would take 5 seconds to sweep. The fastest speed might be 10^{-7} second per division, meaning the trace would complete a sweep of 10 divisions in 10^{-6} seconds. At 10^{-7} seconds per division, a 10MHz sinewave would display as one cycle per division, ten cycles across the screen.

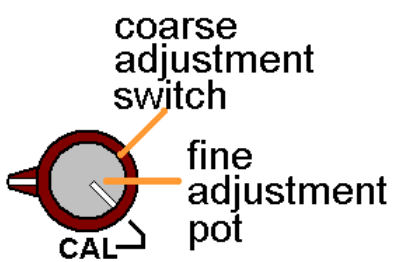


VERTICAL RANGE SWITCHING



Likewise, a scope ought to be able to display waveform voltages from small to large. Waveform height on the screen is controlled by routing the signal through an attenuator / multiplier resistor arrangement like the one in a multimeter, which we studied in EM1. The signal-attenuating switch is calibrated in the same increments as the sweep switch: 1, 2, 5, 10, 20, 50 and so on. When a typical scope is set to maximum sensitivity, a 5mV peak-to-peak signal will occupy one vertical major division. At minimum sensitivity, a 5V P-P signal would occupy the same space.

COARSE AND FINE ADJUSTMENTS



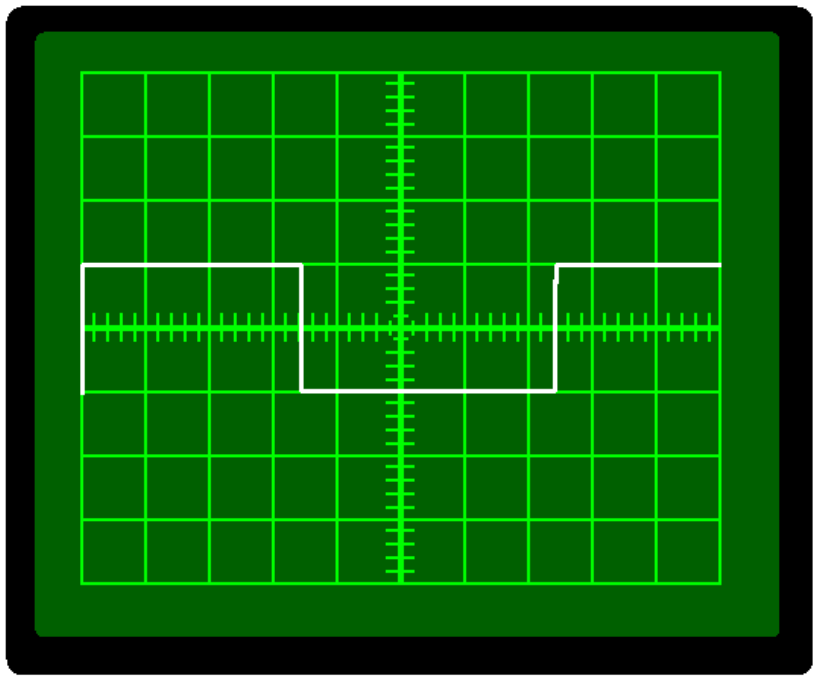
It It is often necessary to adjust the sweep rate or vertical sensitivity more finely than the range switch increments, so scopes are provided with both **coarse** adjustments, to get these parameters in the general ballpark, and **fine** adjustments to trim them. The coarse adjustment is the range switch. The fine adjustment is a pot. Generally, the knob for the pot is located in the center of the range knob, so you've got everything you need in one place.

CALIBRATION

The ability to change vertical sensitivity and horizontal time via pots means you don't really know whether the numbers on the range switches are accurate. If all you want to do is view the waveform, that's no problem. But if you want to count the graduations on the scope face to determine voltage or frequency, those graduations have to be calibrated. So scopes are designed so calibration is accurate when the pots are turned all the way one direction, usually clockwise. When using a scope, keep in mind that the speed or size of the trace on the screen truly corresponds to the range switch setting *only* when the pot is in its calibrated position.

MEASURING FREQUENCY AND VOLTAGE

On the next page is an example of how to use the graduations to determine approximate voltage and frequency.

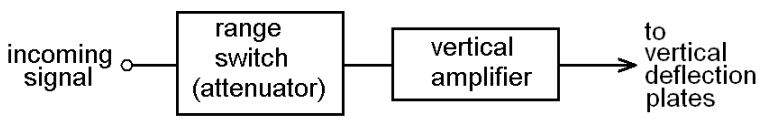


Let's assume we've set the vertical sensitivity to 1V/division and the horizontal sweep to 1mS/division. This means that the distance from top to bottom of each major division, each square, represents one volt and the distance from left to right of each square represents one millisecond. We can see that the voltage of the rectangular wave displayed is about 1V plus and minus, in other words 1V peak or 2V peak-to-peak. The distance between the start of one wave, at the left of the screen, and the next wave, in the middle of the right of the screen, is about 7.4 milliseconds. Since period is the reciprocal of frequency,

$$F = 1 / 7.4 \times 10^{-3} = 137Hz$$ or thereabouts.

VERTICAL AMPLIFIER

This isn't a control, it's a circuit inside the scope. The vertical range switch passes very small signals straight through, while attenuating larger signals. This is how we get signals of any voltage to fit on the screen. However, the signal coming out of the range switch is very small. The next circuit the signal sees is the vertical amplifier, which boosts it to the voltage required to deflect the CRT beam.



Any amplifier has a limit to the voltage it can handle. If you want to enlarge the waveform to look at part of it closely, with the result that other parts of the waveform go off the screen, that's no crime.

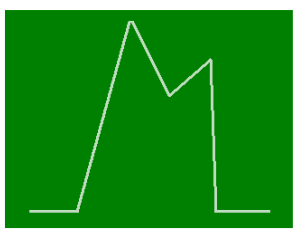
However, you can reach a point where the vertical amplifier is overdriven, and then the strangest things happen. If you're ever doing this and suddenly things go whacky, that's probably the reason. Just a heads-up.

VERTICAL AND HORIZONTAL POSITION

You will often want to move the whole trace up or down or to one side. To begin with, the trace doesn't automatically position itself halfway up the screen, you have to put it there. You'll want to carefully place it right on top of the center line, so when the signal moves it up or down, the displacement can be measured against the division and subdivision lines.

Then again, you may choose to put the trace somewhere else. If you're looking at a waveform that never goes negative, you'll be able to see it more clearly if you move the trace down to one of the lower division lines, then change the vertical range switch to fill the whole screen with the waveform.

centered trace



shifted trace

The circuitry inside a piece of equipment is typically powered by a +DC rail. If you set the trace low on the screen, corresponding to ground which is the lowest voltage in the circuit, and +DC high on the screen, corresponding to the highest voltage in the circuit, every voltage at every point in the circuit can be displayed on the screen.

You may also want to move the horizontal position. This usually results in part of the trace going off the screen to right or left, so you won't be leaving it that way. You move it just long enough to line up part of the waveform with one of the vertical lines, so you can count divisions and calculate period or frequency.

5X SWEEP

When the signal frequency is so high that the waves are munched together on the screen and you can't see them clearly, the natural thing is to turn the sweep rate knob to get a faster sweep. But the sweep frequency only goes so high, which means you can only stretch out the waves so far. This is an unavoidable limit, but the **5X switch** gives us a way to push the limit out a little farther.

When you engage the 5X switch, the horizontal sweep amplitude is increased by a factor of 5, which means 80% of the sweep is off the sides of the screen. In other words, instead of increasing the sweep frequency, the sweep just covers more territory. If there were phosphors on the sides of the CRT, and you could look inside the scope, you'd see that the electron beam is overscanning by a factor of 5. This, of course, makes each part of the trace five times wider, and lets you look at frequencies five times higher, assuming the scope will pass them.

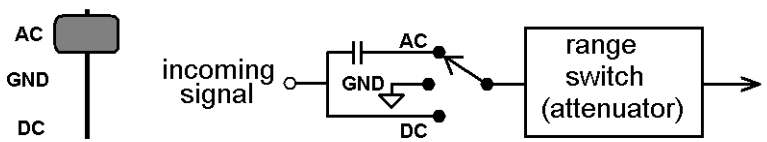
Using the **horizonal position** pot, you can shift the trace back and forth to select which part of the sweep is displayed. Since the sweep is stretched out, it won't be as bright, so you may have to turn up the brightness to see it. Remember to turn it back down again. A super-bright trace can burn a permanent stripe into the phosphor.

Digital scopes have a similar function, but they don't achieve it by overscanning. The number-crunching circuitry selects part of the waveform and changes the resolution, then sends it to the display.

DC/AC/GROUND SWITCH

It is often necessary to examine a small AC voltage riding on a large DC voltage, e.g. when you're following a signal through a defective amplifier to find the point where it's getting lost. The DC voltage will move the trace up or down. If you try to set the range switch so you can see the small AC component, the DC component will push the trace off the screen.

This is the reason for the **DC/AC/GROUND switch**. When set to the DC position, the switch passes the signal straight through to the range switch. When set to AC, it switches in a series capacitor.



The capacitor passes the AC part of the signal but removes the DC component, allowing you to turn the vertical sensitivity up as much as you like. The "ground" position disconnects the signal and connects the range switch to ground. This is helpful in quickly adjusting the vertical position knob to place the trace where you want it.

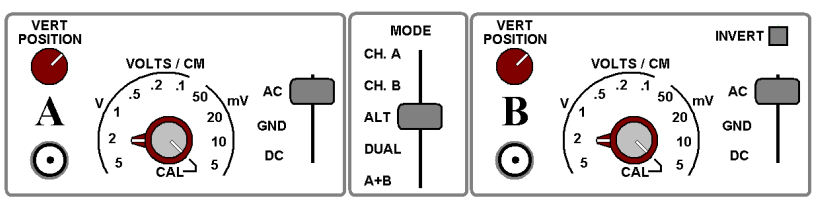
SCALE ILLUMINATION

If you're lucky, the scope will include a scale illumination control that gives you the ability to light up these graduations. This is a nice feature because then you don't have to rely on ambient light to see them. Generally, it's best to place the scope where ambient lighting is low, or where it's in shadow, or to put a dark cowl around the display, because reflections can make it difficult to see the trace clearly.

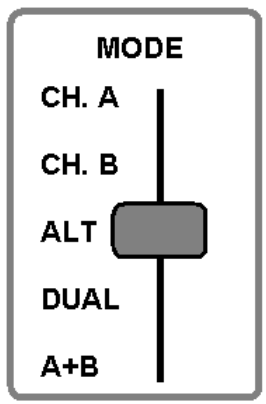
DUAL TRACE SCOPES

All but the most inexpensive scopes can display two separate traces. This is useful in comparing, say, input and output waveforms to see what a piece of equipment is doing to a signal.

The sweep rates are not independently adjustable; the whole point is to compare two waveforms occurring in the same time frame. However, the input attenuators are independently adjustable, so you can simultaneously display a small signal and a large one.



The controls for each input, A and B, are self-explanatory; each trace is provided with the same controls as the trace on a single-trace scope. Each has a vertical position knob, so each trace can be moved up or down on the screen. They can even be superimposed for close comparison. Also, input B has an “invert” pushbutton that flips the polarity of the signal, so when the voltage goes up the trace goes down. We'll come back to that.



The switch in the middle controls how inputs A and B appear on the screen and relate to each other. The first two positions are obvious: one displays only the signal at input A, the other only the signal at input B. It's when we want to display both traces that things get interesting.

The electron beam can't really trace two waveforms at the same time. What happens is that it splits its time between the two. In the **alt** (alternating) position, it trades back and forth, tracing signal A on one sweep and signal B on the next. If the sweep frequency is reasonably high, this looks like two actual traces. The persistence of vision in your eyes, and also the persistence of the CRT phosphor, conceal what's going on.

The user must keep in mind that each signal is invisible at least half the time, while the trace is displaying the other signal. Part of a complex waveform could get lost. When using the alt setting, slow the sweep down so several complete cycles of both signals are shown. That way you can be sure you're seeing everything.

When the sweep rate is slow, alternate scanning doesn't work. The beam is quite obviously trading back and forth, and phosphor persistence doesn't last long enough to fill in the gaps.

In this case, we switch to the **dual** setting. In the dual setting, the beam goes back and forth between A and B during each sweep, hundreds of times per sweep, so both signals are displayed more or less in real time. Some scopes label this the **chop** setting.

The “A+B” setting sums the two waveforms together. This is occasionally useful, but it gets much more useful when signal B is inverted.

When B is inverted, the A+B setting becomes an A-B setting; B is subtracted from A. This allows the user to isolate, for example, small differences between two signals.

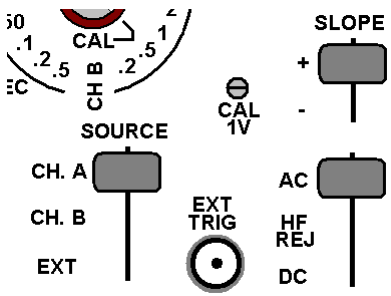
You have to be careful with this setting. The signals are first passed through the vertical amplifiers, then subtracted. The tendency is to turn the sensitivities way up, so you can get a good look at things. But, since the signals are subtracted *after* the vertical amps, turning up the sensitivity usually winds up overdriving them. If things suddenly go haywire when you do this, that's why.

THE DUAL BEAM SCOPE

A very few expensive analog scopes have two electron guns that emit two separate beams. These are “dual beam,” not “dual trace” scopes, and the need for workarounds like alternate or chopped sweeps disappears.

OTHER CONTROLS

Trigger Selector Switches



These controls allow the user to select the signal that the scope uses to derive its trigger. A dual-trace scope will offer a choice between channels, and also an external trigger input. Many signal generators have a trigger output in addition to their regular signal outputs.

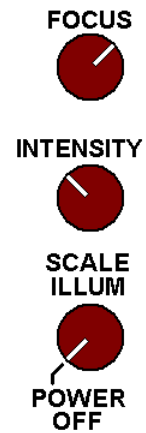
Using this function keeps the scope synchronized with the signal source, no matter how you change the signals or input attenuators. Another switch offers the choice between AC coupling and DC coupling. Setting the trigger controls is more art than science, and talking about it isn't as useful as saying that when you get your hands on a scope, spend some time playing with it.

The HF Reject setting is useful when you're looking at a low to moderate frequency signal. There's all sorts of RF in the air, and if a significant quantity gets into the signal you want to look at, the scope may trigger on the fast-changing RF signal rather than the one you're trying to look at. If you're having trouble getting the scope to trigger properly, that may be the problem. The HF reject setting filters out high frequencies - not from the standpoint of the display, only from the standpoint of triggering.

A third switch offers the choice of triggering when the signal passes a threshold going in the positive direction, or in the negative direction.

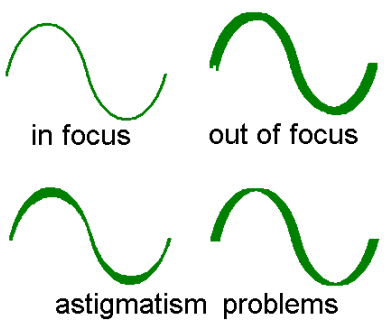
Intensity, Scale Illumination, Focus, Astig, Trace Rotation

The intensity knob adjusts the brightness of the trace on the screen. With analog scopes, if you set the brightness too high, you can burn a permanent scar in the phosphor. This is especially critical when operating the scope in the X-Y mode, discussed below. With digital scopes, phosphor burn isn't a problem; the intensity control is just there for convenience.



The scale illumination control adjusts the brightness of the grid markings. You want them bright enough to see them, but sometimes you have to turn them off to clearly see a trace that falls on top of one of the lines.

Focus, astigmatism and trace rotation apply only to analog scopes. Focus adjusts the sharpness of the dot made by the electron beam. Astigmatism adjusts its relative focus in horizontal and vertical directions. A beam can be in focus horizontally but out of focus vertically, or vice versa, making some parts of the displayed waveform sharper than others. Adjusting the focus will not improve things, because sharpening one part of the wave will defocus another part.



The astigmatism control corrects this problem. Any halfway decent scope should be able to make a thin, sharp trace at every point in the waveform.

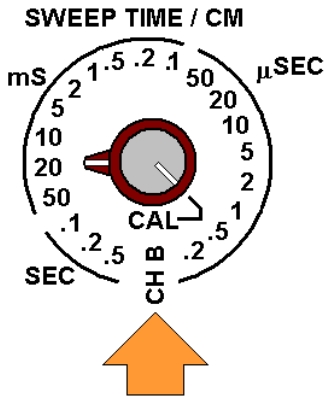
Scale rotation is used to adjust the trace so it's absolutely horizontal on the screen. Stray magnetism anywhere in the vicinity will tilt the trace, and the rotation adjustment is used to correct the problem.

The calibration test point carries an internally generated square wave of precise amplitude, usually one volt P-P. You can couple this to an input to verify that the scope's vertical sensitivity is accurately calibrated. You can also examine the waveform; if it isn't a perfect square wave you know something's funny with the frequency response. This is needed even with digital scopes, which do not drift, because the complete scope system includes the test probes, which can affect results.

THE X-Y SETTING

There's one position on the sweep speed switch that has nothing to do with sweep speed. In the **channel B** setting, the beam doesn't sweep at all. Instead, the B-channel signal is used to deflect the beam horizontally, while the A-channel signal deflects it vertically. In the absence of signal, the beam sits in the center, making a glowing dot.

(Turn down the intensity when using this setting.) This turns the scope into a graph-drawing machine that displays A as a function of B. Let me give a few examples to show how useful this can be.



In figure 1 at lower right, the same signal is connected to A and B inputs. The display is a line tilted upwards at 45 degrees, as you'd expect; for every CM of horizontal movement, there's an equal CM of vertical movement.

In figure 2, the same signal as before is connected to input B, but the signal goes through an amplifier on its way to A. The flat top and bottom shows that the amp is beginning to clip.

In figure 3, the amp is beginning to overload, but the distortion comes on gradually. Looking at a sine wave, you'd never see this; gentle distortion of a sine wave must reach about 10% before it is obvious.

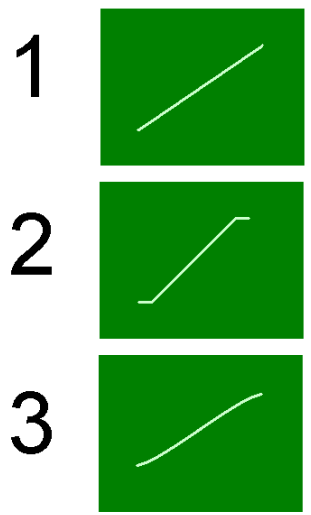
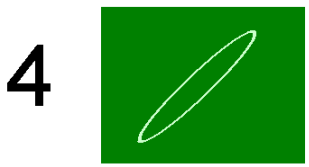
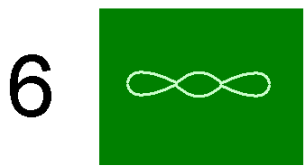
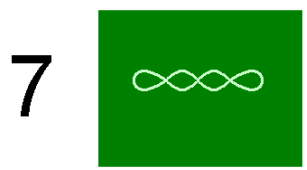


Figure 4 shows a phase difference between the signal at A and at B. If you were adjusting the relative phase of two sine waves, you could feed one to the A input and one to the B input, then vary the phase of one of the signals until the loop closed up. This works best with high-quality scopes; inexpensive scopes often have some built-in phase difference between horizontal and vertical circuitry, thus can't give you a reliable phase indication at high frequencies.





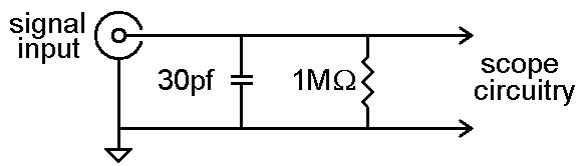


Figures 5, 6 and 7 show that the frequency of the sine wave at A is twice, three and four times the frequency of the sine wave at B. These are called **Lissajous patterns**.

A single-trace scope usually has a separate horizontal input for the X-Y setting. This isn't quite as good, but it works. The advantage of a dual trace scope is that the X-axis deflection has the same range of control as the Y-axis, with the same calibration.

INPUT LOADING

As with a VOM, the user must always keep in mind that the act of attaching a measuring device may upset the measurement. The inputs to an oscilloscope usually have about a 1 megohm resistance in parallel with as much as 30pf of capacitance. The resistance is fairly high, but capacitive reactance falls below 1MΩ already at 5kHz. When the signal frequency reaches 5MHz, this input is essentially a 1kΩ capacitive load, which is a pretty heavy load to be attaching to the innards of a circuit.

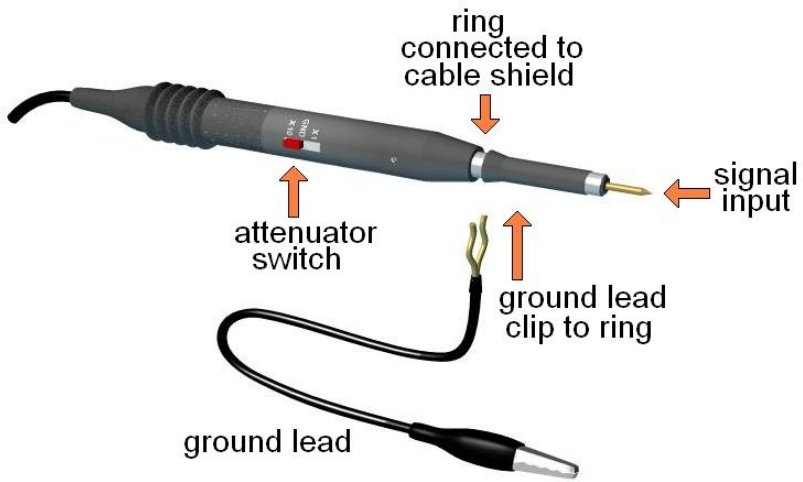


Moreover, measurements are taken via probes attached to the inputs, and that adds more capacitance. A typical probe connected to the scope via a 3-foot cable adds 100pf. Better probes, using special low-cap cable, have less capacitance.

Bottom line: simple probes aren't good for measuring frequencies over a few hundred kHz, unless you're looking at a point in the circuit with very low impedance.

ATTENUATOR PROBES

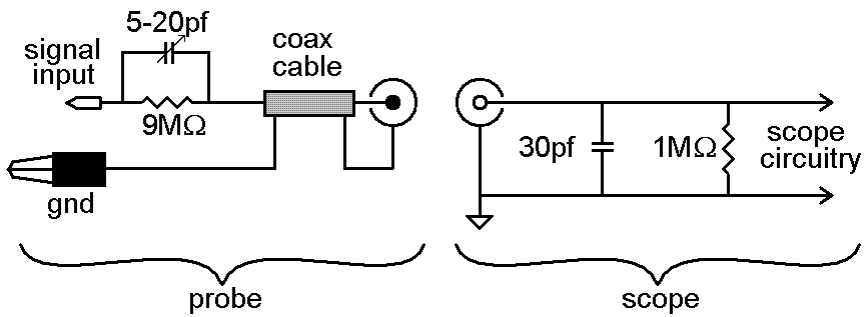
An attenuator probe has a switch that allows you to insert a 10:1 voltage attenuator, right in the probe. This has two uses: reducing the load on the circuit you're measuring, and increasing the maximum voltage the scope can handle.



If we go back to the input attenuator of our representative scope, we see that the least sensitive setting is 5V/CM. When the switch is rotated to this setting, a 5V change at the input will move the trace one major division up or down. Since there are eight major vertical divisions, this means the largest waveform that can be shown is 5 x 8 = 40 volts P-P. This can be increased by rotating the "fine" knob, but then we lose the calibration.

Part of the reason scopes aren't made to handle greater input voltages is that it isn't a good idea for those voltages to appear on the input connector, which is usually a "BNC" type that's an excellent connector from a signal-handling standpoint, but isn't designed for high voltages.

The proper way to display a high voltage is to switch in the 10:1 attenuator in the probe. This inserts a 9MΩ multiplier resistor and small capacitor in series with the input, creating a voltage divider that knocks the signal down before it even gets to the shielded cable. This converts the 5V/CM range switch setting to 50V/CM, allowing the scope to display a signal up to 400V P-P. Using a scope this way requires some presence of mind on the part of the user; you have to remember which position the probe switch is in, or your measurements will seem to be off by a factor of 10:1.



The other thing that happens when you switch in the 10:1 attenuator is that the input impedance goes way up, which is to say the loading goes way down. In the 10:1 setting, the multiplier resistor is paralleled by a small variable capacitor that forms a capacitive voltage divider with the input and cable. Typically, the final capacitance is about 12pf. Deluxe probes have less.

Many techs get in the habit of leaving the probe in the 10X position, and mentally multiplying the input attenuator setting by 10. They only switch to 1X when looking at a signal so small the scope wouldn't display it with the attenuator switched in.

THE GROUND LEAD

I've mentioned this before, and I'll talk a lot more about it in chapter 10. It's not good enough to just get the scope input connected to the unit you're testing. You also have to get the scope ground connected to the unit's ground. And, as you'll see in chapter 10, "ground" is a slippery concept.

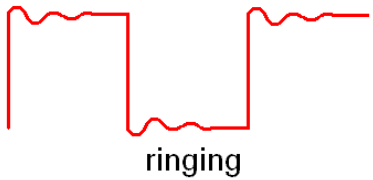
A test probe has a ring near the tip that's connected to the coax cable shield. The probe comes with a short lead that attaches to this ring, with an alligator clip at the other end to attach to ground. This is the ground lead you're supposed to use to attach to the chassis of the unit you're working on.

You're going to break that rule a million times. The ground lead is often too short to allow the probe to reach the point in the circuit where it has to reach. You'll find yourself using another clip lead to extend the ground, and many times, in frustration, you'll just run a “wild” ground wire from the scope itself – there's almost always a ground terminal somewhere – to the unit under test. Problem fixed.

But there's a reason the supplied ground lead is so short, and you should know what it is so you'll know how bypassing it might affect your measurements.

At low frequencies, one end of a grounded wire is the same as the other end. But at high frequencies, there's a time difference between the ends of the wires. What this means is that, at 100MHz, one end of a 4-foot test lead isn't firmly connected to the other end. It isn't so much a cable as a tunnel, a conveyance that guides waves along a path. For this to work, the ground – no, it isn't a ground; the *reference* must travel right along with the signal. If the reference goes from unit under test to scope by any other path, things can get screwed up.

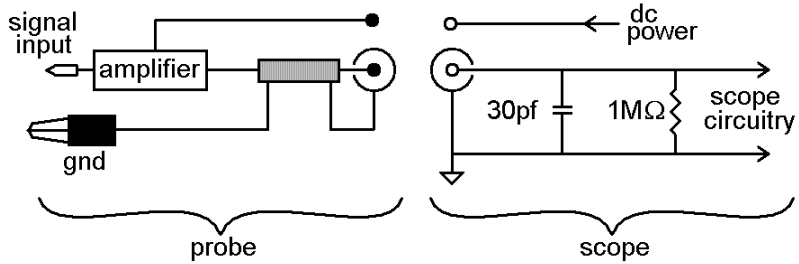
For example, in a square wave, you can get **ringing**, which is related to the resonant behavior of a bell. The waveform you're looking at may actually be ringing, or it may be an artifact caused by careless grounding. It's important to know which it is.



I'm not saying you should never bypass the “system." I myself am extremely careless with scope grounding, but then, I usually work with frequencies under 50kHz. At 100MHz, and especially with digital signals which are not sinewaves, the waveform on the scope will bear no resemblance to what's going on in the circuit unless you treat the test lead as a transmission system, not just a wire. If you're testing something, and you haven't been careful with the grounds, and the waveforms seem to be full of strange glitches, clean up the grounds and look again.

ACTIVE PROBES

At frequencies of hundreds of megahertz, even the small capacitive load of a 10X probe is a problem. High-speed scopes isolate the test point from the cable capacitance by building amplifiers right into the test probes. The amplifier, not the input signal, supplies the current necessary to drive the cable and scope input



An active probe can have a capacitance as low as 2pf. You have to be careful with it; if you break one, you'll have to get another identical one from the scope manufacturer. To my knowledge, nobody makes a generic amplified probe that will work with a wide range of scopes. And active probes are not cheap.

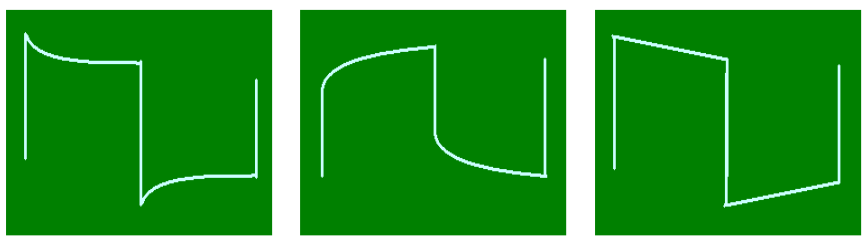
The easiest way to damage an active probe is to connect it to a voltage that's higher than what it's designed to handle. Remember, the scope's range switch comes after the probe, so turning it down doesn't reduce the voltage that the probe's amplifier sees. Read the owner's manual so you know its limits.

Chances are, you won't be dealing with active probes early in your career. If that time does come, it would be a good idea to get a little specialized training.

CALIBRATION AND SCOPE LIMITS

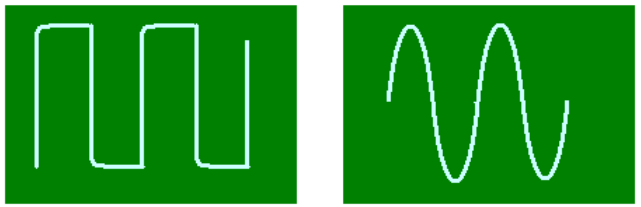
The biggest mistake people make when using test gear is to assume what they see is automatically true. It can't be said often enough, and though it applies with special force to oscilloscopes, it also applies to all other test gear: you need to become so intimate with it that if you cheat on your gf/bf it'll know. Wizard-level techs even study the internal circuitry. You'd be surprised how often that comes in handy.

A scope is not really ready for use until you've calibrated it, checked its frequency linearity, adjusted it so square waves have perfectly square corners, and adjusted the probes so waveforms look the same when set to 10X as when they're set to 1x.



excessive high frequency response | deficient high frequency response | deficient low frequency response

Then, you need to become familiar with the scope's limits. For example, if your scope won't display anything over 15 MHz, a 10MHz square wave is going to look like a sine wave. That square wave's lowest harmonic is 30MHz



10MHz square wave on a 50MHz scope | 10MHz square wave on a 15MHz scope

(square waves don't have even numbered harmonics), so the fundamental is the only thing that will get through.

You also need to get a good feel for apparent anomalies that are natural results of the measurement process. Suppose you're looking at a low frequency square wave with the input switch set to AC, and you see some tilt. That's what happens when you put a capacitor in series, it's normal.

Does it look like there's some DC on the input, when there isn't supposed to be? Switch it to GND and check to be sure you haven't accidentally bumped the vert position knob.

Age isn't the only thing that makes techs go bald. They also pull their hair out, unable to locate a nagging problem in a unit they're servicing because the problem isn't in that unit, it's in the way they're using the test gear. So now you've been alerted.

STORAGE OSCILLOSCOPES

It's often handy to have a scope that can store a waveform, e.g. to catch a glitch that only occurs once in a while. There were analog storage scopes, but digital scopes run rings around them. Anyone who needs to capture and store waveforms for later analysis needs a digital scope, period.

DIGITAL SCOPES

Digital scopes are essentially computers, and their displays are small LCD monitors. They do not “scan” a trace onto the screen. The scope captures the signal waveform, converts it to digital form, finds the zero-crossings, uses those to anchor the waveform, puts a copy in memory, and then presents it to the monitor as a kind of photograph.

Obvious differences from analog scopes:

1. The focus, astigmatism, and beam rotation controls disappear. With digital scopes, there's nothing to go out of adjustment.
2. A lot of the switching and adjustment is handled electronically, with buttons instead of knobs and switches.
3. Triggering functions are improved.
4. Since the signal waveform is stored in digital form, it's easy for the scope to retrieve it from memory and put it on the screen. This allows you to capture and examine signals to catch occasional glitches.
5. All but the cheapest digital scopes offer extensive analysis tools that go way beyond showing the waveform. For example, most will analyze and graph the spectrum of a waveform. You can get this function in scopes starting well under $1,000.

BUYING AN OSCILLOSCOPE

Your employer will provide you with a scope to use on the job if that's what the job requires, but God bless the child who's got his/her own. You don't have to spend a fortune for the hottest, latest thing. The very best scopes made 20 years ago were so far ahead of their time that they hold their own even today, at a tiny fraction of their original prices.

Here, for example, is a Tektronix 475A, which was a high-end general purpose scope in the 1980's. It's considered a clunky museum piece today.

And yet...it has a dual trace, 275MHz bandwidth, which is still pretty deluxe, and a built-in multimeter on top that's internally connected to the signal input, so you can measure the signals you're looking at without running a rat's nest of wires. Note the handle. This beast weighs around 40 pounds, which was “portable” by 1980's standards. But as a bench scope, it will serve most techs as well as a new one. This is the one I got for $300. Many techs have no reason to want more.

With excellent scopes like this available so reasonably, there's no reason for a tech to put up with junk. A scope of your own gives you something valuable: the freedom to fiddle with it on your own time and really master its controls, functions and capabilities.

High-test scope makers like H/P (now Agilent) and Tektronix made modular systems consisting of 19-inch rack boxes that contain a scope and slots for other kinds of test gear made by the same manufacturer – meters, generators, analyzers and so on. The one shown here has two signal generators, a frequency counter and a multimeter, all top quality.

Some scopes include voltage and frequency counters that display right on the screen, alongside the waveform trace. They may have cursors you can move across the trace to allow you to measure instantaneous voltage at any point in the waveform.

Rules of thumb:

1. Don't buy a used scope without a 30-day guarantee. If it isn't working when it arrives, it's an expensive doorstop.
2. Don't spend over $400 on an analog scope. Digital scopes with analytical capabilities no analog scope can match start below $1,000.
3. If you need small and battery operated, there are some impressive digital scopes available, but check the capabilities carefully. Many are severely limited in what they'll display. If you're willing to settle for an analog battery-operated scope, you can get one for as little as $100.
4. If you're planning a career path that involves a lot of digital troubleshooting, don't even consider an analog scope. You need the digital scope's storage capabilities.
5. Don't buy anything at all yet. Work in the field for a while until you know what you want. Keep your eye on Ebay.

OTHER TEST GEAR

I've devoted most of this chapter to oscilloscopes, but there are some other pieces of AC test gear you should know about.

THE SPECTRUM ANALYZER

You've already seen X-Y graphs where the X-axis corresponds not to time but to frequency. This is the kind of chart used to display the spectral content of a signal. It can also be used to display the frequency content of, for example, the signal being picked up by an antenna.

We could create a graph where the X-axis covers the band from 540 kHz to 1600kHz, with a bunch of vertical lines whose positions and heights represent the frequency and signal strength of all the AM radio stations on the air in a particular area. The machine that does this automatically, in real time, is the **spectrum analyzer**. It is an oscilloscope married to a fairly complex bunch of circuitry that sorts out the energy at its input by frequency and strength.

A spectrum analyzer can be used to:

- display all the transmitters on the air over a given band of frequencies
- find an open channel
- analyze the purity or impurity of a sine wave
- determine whether a transmitter is putting out a clean signal
- determine whether an antenna is messing that signal up
- help design a filter to reduce the electrical noise radiation of a machine
- ...and a thousand other things.

I could say a lot more, but spectrum analyzers are all different. Some are designed to analyze only a certain range of frequencies; some are designed to cover an extremely wide range. Most of them are digital nowadays, and that comes with traps for the unwary. If you wind up working with spectrum analyzers a lot, you'll get special training.

FREQUENCY COUNTERS

A frequency counter derives a trigger voltage from a signal waveform, just like a scope. But, instead of using it to start a sweep, the trigger increments an electronic counter.

The counter totals up the number of trigger events during a specific interval such as a second, then transfers that count to a visible display.

Getting a proper count is sometimes tricky. If the counter is looking at a waveform that crosses zero more than once per cycle, it will count each crossing separately. Also, if the signal has some contamination on it, e.g. an audio signal that has picked up some RF, the counter will count the crud along with the desired signal.

A counter is most useful in verifying that a frequency is exactly what it should be; you should have a general idea of the correct number so you'll know if the displayed number is way off. Better counters have threshold and filter controls that assist you in getting proper trigger behavior.

One more word about digital scopes: many of them perform frequency counter and spectrum analysis functions. The higher cost of a digital scope may be offset by the fact that you then don't need to buy those other pieces of test gear.

DISTORTION ANALYZERS

One of the Great Truths about amplifiers, transmitters, receivers and other kinds of circuits is that they always modify the signal they're working with, at least a little. This is what techs call **distortion**.

Whether or not a particular kind and amount of distortion matters depends on the application, but when it does, we need to have ways of sniffing it out and analyzing it.

Sine wave distortion has to get pretty gross before it's visible on a scope. That gently curving form doesn't give us any signposts to latch onto; it can be squished, stretched and bent a fair amount before we can tell by looking. So various kinds of distortion analysis equipment has been invented that looks for specific markers of specific kinds of distortion. Why do we need to know? Distortion:

- makes amplifiers sound crummy
- creates unwanted signals at various frequencies
- creates radio interference
- screws up video images
- and so on.

Distortion is usually measured in percentages. The best analyzers also include a scope, or have an output that can be fed to a scope, where the distortion is strained off from the desired signal and presented separately, in all its naked ugliness. This can be of immense help in finding the cause and fixing the problem.

CAPACITOR- INDUCTOR TESTER

This is the equivalent of an ohmmeter for reactance. In its simplest form, it gives you a pair of leads that you can clip across a capacitor or inductor and read the value. More complex versions permit you to sweep a frequency band, which helps you investigate the characteristics of cables, circuit components and so on. A good tester will also tell you if there's a partial short or open circuit.

INDUCTIVE AMMETERS

With AC we can do something we couldn't do with DC, at least at low frequencies: measure current without breaking into the circuit. Most people know this kind of meter as an "Amprobe," though that's a registered trademark. Electrically, it's just a current transformer (see chapter 7), but the core is split and can be opened, placed around a wire, and closed. For this to work, you have to clamp around a single wire. If you clamp around a cable with two wires, with current flowing from the supply on one wire and returning on the other, the fields will cancel and you'll get no reading. Also, you will get no reading if the current is DC.

SUBSTITUTION BOXES

In EM1, we looked at a resistor substitution box. Capacitor and Inductor substitution boxes may be useful to you in the same ways, depending on what kind of work you wind up doing. You can make your own resistor boxes, but it's best to use commercially made L/C boxes because it's too easy to accidentally build in some stray L or C and wind up with a box that doesn't serve its intended function.

There are hundreds of other kinds of test gear designed to examine various signals in various ways. We won't examine them because they're built for specialized applications. Next we'll look at equipment that makes signals for those analyzers to analyze.

SIGNAL GENERATORS

The requirements for a signal generator are:

- The signal must be clean enough and accurate enough that any fault observed while testing can be assumed to be in the unit being tested, not the generator.
- The output should be stable, without drift
- The output should not be overly sensitive to loading, and certainly not to the point where it gives rise to measurement errors. In practice this means it should either have a “stiff” low-impedance output or a precise, specified impedance that the user can take into account when making measurements.

We're going to look at only the most common and broadly useful types of generators; special-purpose ones are far too numerous to list. I'm only going to hint at the technology involved, because you haven't studied it yet – that's coming up in chapters 5 thru 9.

A circuit that generates a repeating waveform is called an **oscillator**. A signal generator is an oscillator plus adjustment circuitry, plus an amplifier between the oscillator and the load to keep loading changes from upsetting the oscillator's operation.

SINE WAVE GENERATORS

The most basic signal generator creates a sine wave. Generators differ mainly in the ways the signal is generated, the purity of the sine wave, and the frequency range covered.

RF Sine Wave Generators

An RF generator is usually based on an **LC tank**, a tuned circuit made of an inductor and a capacitor. Chapter 9 covers LC tanks in detail. The waveform is naturally a sine wave, so no special tricks have to be performed to get a pure output. The range of frequencies available over a single turn of the dial is usually around 3:1. A range switch permits the generator to cover an overall range from perhaps 100KHz to 100MHz.

AF Sine Wave Generators

AF generators are based on resistance-capacitance or **RC** circuits.

A single turn of the knob covers about a 10:1 frequency range. An AF generator has to cover an overall range of 20Hz- 20kHz to be any use at all; better ones go 5Hz to 100kHz.

In both cases, the RC or LC circuits have a natural **resonant** frequency, so they ring like a bell. Of course, if you strike a bell, the sound eventually dies away. An oscillator generates a continuous wave by feeding just enough energy into the resonant circuit, taken from its own output, to keep the oscillations going.

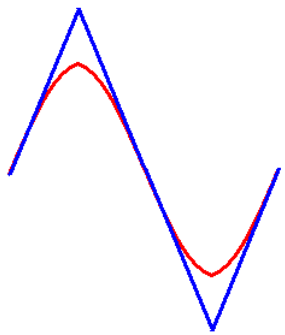
The greatest strength of LC and RC sinewave generators is spectral purity. The weakness of the RC generator is signal level stability. AF generators in particular need sophisticated circuits to keep the sine wave from growing until the circuitry clips, or dying away. Not an insurmountable problem, though. It'll have more to say when we study oscillators in EM3.

THE FUNCTION GENERATOR

The function generator creates a variety of waveforms – at a minimum, square, triangle and sine waves. Some also make sawtooth and rectangular waves. There is usually an input to allow an external source to modify the frequency or level of the signal, and an output to independently trigger an oscilloscope.

There are no tuned circuits in a function generator; the waveforms are generated by timing circuits. This works fine for square, triangle and sawtooth waves, but generating a sine wave without a tuned circuit is difficult. A function generator does it by taking a triangle wave and squashing the peaks.

This gives a reasonable imitation of a sine wave, but it isn't harmonically pure. Few function generators are able to create a sine wave with less than 1% distortion. If you need to measure spectral purity, a function generator is just the wrong tool to use.



SWEEP GENERATORS

A sweep generator creates a waveform, usually a sine wave, that automatically and continually sweeps over a large frequency range. A 10:1 range is common; a 1,000:1 range is achievable. The generator will have a separate connection that supplies a sawtooth wave to an oscilloscope, so an entire frequency sweep can be displayed. This could be called a poor man's spectrum analyzer, and it's every bit as good in many applications.

There are essentially two kinds of analog sweep generators. One, called a **beat frequency oscillator**, uses the difference frequency between two high-frequency oscillators to derive the output frequency. This is a basic part of radio technology, and we'll cover it in detail in EM6. The other kind of sweep generator is a swept function generator. There are also newer ways to do this digitally, which we'll look at in chapter 4.

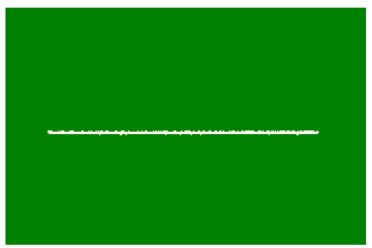
Some sweep generators are built to sweep a fairly narrow band that corresponds to a single AM, FM or TV channel. This permits quick adjustment of the complicated system of tuned circuits in a receiver, or verification that the response of a system is uniform over a specified band of frequencies.

NOISE GENERATORS

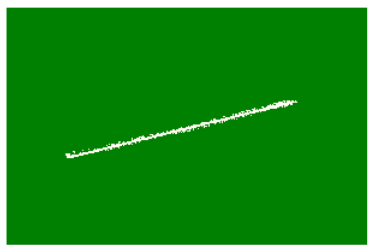
In chapter 3, we defined noise as a randomly varying voltage that has energy all across the frequency spectrum. This implies that, between any two frequencies, there will be a specific amount of energy, and the farther apart those two frequencies are – the wider the band between them – the more noise energy it will contain. There are two specific kinds of noise that correspond to two distributions of energy.

White noise is “natural” noise, found in nature. White noise has an equal amount of energy in bands of equal width; e.g. there is as much energy between 100Hz and 200Hz as between 10,100Hz and 10,200Hz. If you feed this into a spectrum analyzer with a linear frequency scale, it will display a flat line across the spectrum, showing that there's equal energy in every frequency band.

However, if the same signal is fed to a spectrum analyzer with a proportional or logarithmic frequency scale, the line will tilt upwards.



white noise on a linear spectrum analyzer display



white noise on a logarithmic spectrum analyzer display

This happens because, if we're measuring frequency in proportional increments, every increment will be twice as wide as the previous one, so there will be twice as much noise energy in it.

This problem is corrected by passing the noise through a filter with a complementary tilt; that is, one in which attenuation increases as frequency increases.

Noise filtered in this fashion is called **pink noise**, and instead of having equal energy per absolute frequency increment, it has equal energy per proportional frequency increment.

Some noise generators are **pseudo-random**. They follow a set, repeating pattern, so they are not actually random. However, they contain energy across the spectrum and are functionally equivalent to true random noise generators for the purposes of testing. Their advantage is that the output level is automatically 100% predictable and stable.

We've only covered a small part of the total variety of test gear available. In EM5, we're going to spend a lot of time on service procedures and malfunction-finding techniques, and then we will go into more detail about how all this cool test gear is used.

End of EM2 chapter 4

EM2 Chapter 5
AC COMPONENTS

After completing this chapter, you will be able to:

- *identify capacitor types, construction and uses*
- *describe the construction of variable capacitors*
- *identify inductor types, construction and uses*
- *describe the construction of variable inductors*
- *identify transformer types, construction and uses*
- *describe capacitive and inductive reactance*

In EM1, we looked at a number of components that are used in DC circuits. In this chapter, we will study several additional components that are only used in AC circuits, or components we've already studied that have additional uses in AC.

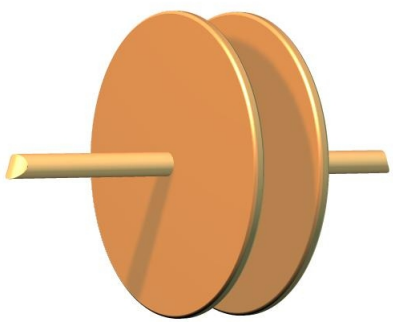
CAPACITORS

A capacitor is so named because of its capacity to hold an electrical charge, an ability called **capacitance**, or **C**. Capacitance is measured in **Farads**, abbreviated **F**, after a scientist named Michael Faraday.

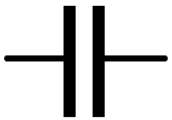
Because a Farad is a very large unit of storage, most capacitors are measured in **microfarads**, abbreviated **µF**, **µFD** or **µfd**, 10^{-6} Farad. Sometimes you will see a capacitor rated in **MF**. That's microfarads, despite the incorrect multiplier. Capacitors are never rated in millifarads or megafarads. Much smaller capacitors are measured in **nanofarads**, **nF**, $^{10-9}$ Farad, or **picofarads**, **pF**, 10^{-12} Farad.

Capacitor Construction

Any object can be charged, of course, but a capacitor stores a charge *difference* – a potential – between two terminals. A capacitor consists of two conductive surfaces with an insulating **dielectric** between them. The simplest capacitor is two metal **plates** placed close to each other, with air as the dielectric. If electrons are pulled out of one of the plates and pushed onto the other, that creates a voltage across the plates.



In the U.S. the schematic symbol for a capacitor is a pair of straight, heavy lines representing the plates, with a wire extending from either side. Note that the heavy lines are equal length, as opposed to the symbol for a battery where one line is longer than the other. That's how you tell a battery symbol from a capacitor symbol.


capacitor

It is much easier to store a significant charge in a capacitor than to put an equivalent charge on an isolated object. Because the plates are close to each other, the electrons on the negative plate “feel” the attraction of the protons on the positive plate, which cancels much of the repulsion from nearby electrons. Also, the dielectric itself stores energy, as its electrons and protons change position slightly in response to the electric field that exists between the plates.

If you've taken high school physics, you've studied “potential” and “kinetic” energy. Kinetic energy is the energy of movement, of momentum, like the energy in a moving bullet. It comes from the same Greek root as our word “cinema,” pictures that move. Potential energy is stationary energy, like the energy in a tensed spring. Capacitors are potential energy storage devices.

The size of the charge on a capacitor per volt of potential depends on three things:

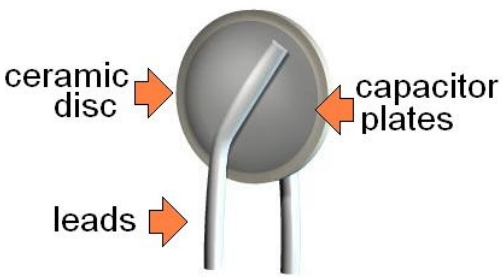
- the size of the plates,
- the distance between the plates, and
- the nature of the insulating dielectric between them.

Let's look at these factors one at a time.

Size of Plates

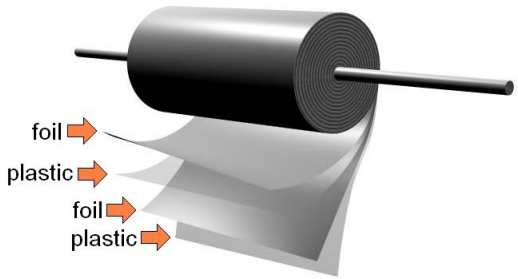
Obviously, the larger the plates, the larger the charge a capacitor will store at a given voltage. If the area of the plates is doubled, the charge per volt will double, because twice as many electrons can be pushed in. If one plate is much larger than the other, or if the plates are near each other over only part of their area, the parts where they do not overlap do not contribute significantly to the capacitance.

Very small capacitors use a ceramic material, or mica, or some other solid insulator for the dielectric. The metal plates then take the form of a plating on either side of this dielectric.



disc capacitor with encapsulating layer removed

Where more capacitance is desired, it is possible to make the plates a lot larger without making the capacitor physically impractical by making the plates out of sheets of foil and rolling them up. In this case, the dielectric is two insulating sheets rolled up between the plates. Old capacitors used paper insulation; newer ones use various types of plastic film. There's another advantage to this kind of construction: both sides of a rolled plate are near a plate of opposite polarity, thus doubling the capacitance. In this kind of capacitor, the plates are connected to wire leads coming out opposite ends. It is often useful to know which lead is connected to the outermost foil layer; this will be indicated by a stripe on the capacitor near one end.



Spacing

The closer the plates are to each other, the more charge the capacitor will store per volt. This is a two-edged sword because, if the voltage gets too high, a spark will jump between the plates, and the closer the plates are the less voltage this takes. So a capacitor with really close plates will store a comparatively high charge per volt, but won't handle a high voltage.

Dielectric

The nature of the insulating material between the places, which is called the called the **dielectric**, influences the amount of charge per volt a capacitor will store. It also determines how close the plates can be, because different materials have different voltage breakdown limits.

The insulating properties of a material are called the **dielectric constant**. Air and space have a dielectric constant of 1. Other dielectrics and their constants:

paper	2.5-3.5
quartz	4.5
poly plastics	2-3
mica	5-6
glass	4-7
ceramic	50-15,000 depending on type

A higher dielectric constant means that a pair of plates spaced “X” distance apart can store that many times more charge per volt than an air-gap capacitor.

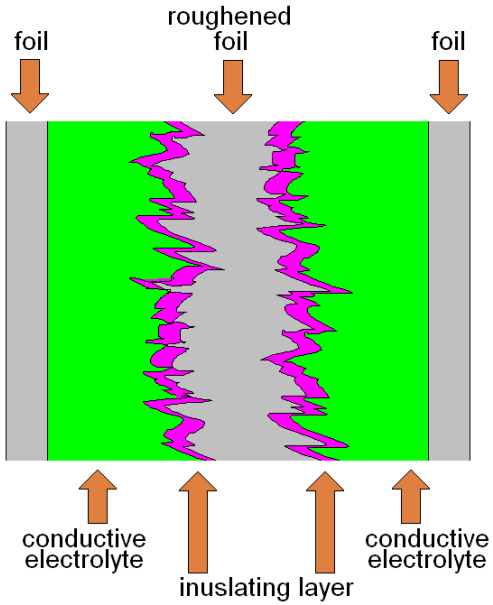
If the voltage limit of an air dielectric capacitor is exceeded, the only thing that will happen is that a spark will jump and the capacitor will be discharged. Some forms of lightning protector are essentially capacitors with one side connected to earth ground; sparking across the dielectric is exactly what is supposed to happen. On the other hand, a solid-dielectric capacitor is likely to be destroyed if the voltage limit is exceeded, because the spark will punch a hole in the dielectric. This is called **breakdown** of the dielectric and usually means the capacitor must be replaced.

The capacitors studied so far are called **dry** capacitors. They are **non-polar**, which means that they will store a charge in either direction, so it doesn't matter which way they are connected.

Electrolytic Capacitors

There is another class of capacitors called **electrolytic** capacitors. In an electrolytic capacitor, one of the foil plates is etched to roughen it. A rough surface has a much greater area than a smooth one. This surface is coated with a thin layer of dielectric material. Then the foil plates are rolled up and put into the capacitor case, and the capacitor is filled with electrolyte, which is an electrically conductive liquid. This liquid is in contact with, and electrically connected to, the plate that was not etched.

Thus, when we talk about the space between the plates in an electrolytic capacitor, we are talking about just the distance between the etched plate and the electrolyte, which is only the microscopic thickness of the dielectric layer. So we have both a greatly increased surface area and a greatly decreased dielectric thickness.



This permits an electrolytic capacitor to have much greater capacitance than a dry capacitor of the same size.

But we have to be careful. Applying a DC voltage of the correct polarity across the capacitor does not harm the dielectric, and in fact keeps it healthy. But a reverse voltage will dissolve the dielectric, and then the capacitor is ruined. Thus, an electrolytic capacitor is **polarized** and you have to pay attention to which way you put it in a circuit.

A nonpolar electrolytic can be faked by connecting two regular electrolytics back to back, so that no matter how the voltage appears across the pair, one or the other will be correctly polarized and will block current from ruining the other one. You can buy a nonpolar electrolytic, which is two capacitors in one case. Or, you can take two capacitors and make one yourself.

In an electrolytic capacitor that isn't used for a long time, the dielectric may deteriorate. There is a procedure for reforming a capacitor which you can find on the web. However, they're cheap enough nowadays that it's usually easier to just buy a new one. If the dielectric has dried out, you have no choice. If you're working on a piece of equipment that's decades old, you will probably have to replace all the electrolytics.

Oil-Filled Capacitors

Some high-power, fairly high-voltage capacitors are made in the form of a dry capacitor that is anything but dry – it's a sealed metal can where the foil plates and dielectric film are bathed in oil. The oil serves both as additional dielectric and heat conductor, moving heat from the working parts to the can where it can be dissipated into the air. Oil capacitors are used where high AC currents are present. They are nonpolar, but like electrolytics, if they spring a leak and dry out, they become useless.

SUPER- AND ULTRA-CAPACITORS

Recently, a class of capacitors has been developed whose storage capacity is much greater than even an electrolytic capacitor. For the first time, it is possible to make a capacitor of several Farads in the space of a bottle cap. **Supercapacitors** are beginning to replace batteries in many applications. They can operate over a wider temperature range and take many more charge/discharge cycles than even the best batteries. At the moment, supercapacitors can only handle a few volts, but this may change.

Capacitor Leads

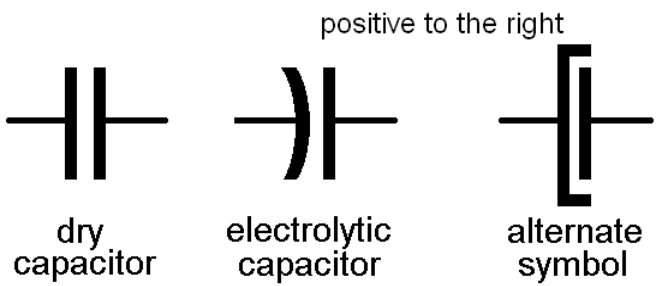
A tube-shaped capacitor may have one lead coming out the middle of either end. These leads are located on the axis of the tube, and are called **axial-lead** capacitors. Or, both leads may come out one end. These are called **radial-lead** capacitors and they're intended to stand up on a PC board. Even if a capacitor is some other shape – a disc for example – if it has both leads coming out one side it's usually called a radial capacitor.

axial leads

radial leads

SCHEMATIC SYMBOLS

As noted above, in the U.S. the schematic symbol for a dry capacitor is a pair of straight lines. The symbol for an electrolytic capacitor is one straight, heavy line and one curved one. The straight line represents the positive terminal. Alternately, the symbol may take the form of a straight line surrounded by a C-shaped line; again, the straight line is positive.

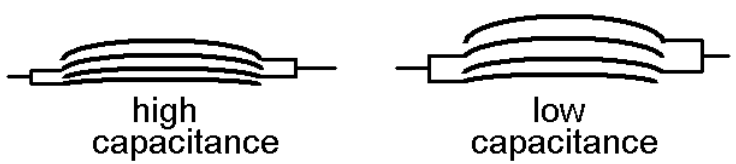


Not everyone follows this convention. Dry capacitors are often shown with one curved line, and schematics of European equipment often show dry capacitors with a C-shaped line. These are not necessarily mistakes. As mentioned, even with dry capacitors, it sometimes makes a difference which way they are connected because, when the foil strips are rolled up, one will be on the outside, and this is sometimes important. More on this in EM5, when we get to servicing.

VARIABLE CAPACITORS

If a capacitor is made so the plates can be moved, the capacitance can be adjusted. The two most common ways of doing this are to move the plates closer or farther from each other, and to slide or rotate them sideways so part of the surface of each plate does not have a plate of the opposite polarity near it. The choice usually depends on how the variable capacitor is going to be used.

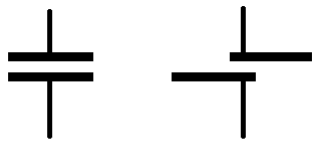
If the purpose is to make a one-time adjustment that will seldom or never have to be changed, the adjusting means will be a screw head. The capacitor usually has many plates in a stack, with insulating sheets between metal layers. The metal layers are cupped a bit, to give them some spring, so they will naturally be spaced apart. In this state the capacitance is lowest. Then, if the screw is tightened, the plates are squeezed together and the capacitance increases.



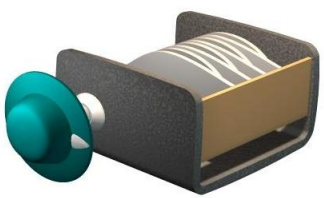
This type of capacitor is called a **trimmer** capacitor, because it's usually connected in parallel with a much larger capacitor and used to “trim” the total capacitance to a precise value, like the trimmer resistors we studied in EM1.

If the purpose is to make frequent adjustments, the capacitor will still be made with a stack of plates, but every other plate will be attached to a shaft that can be rotated.

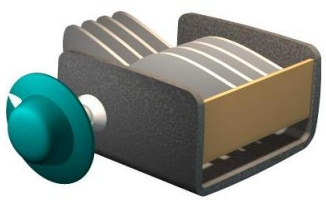
This allows the plates' area of overlap to be changed. In this sense it is like the rotary rheostats we studied in EM1. A capacitance variation of 10:1 can easily be achieved. Often a variable capacitor will include both types: a rotating part used to vary the capacitance, and a screw-adjusted part to trim the exact capacitance at the rotation limit.



high capacitance | low capacitance

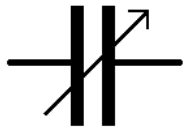


variable capacitor at maximum



variable capacitor at minimum

The schematic symbol for a variable capacitor is a regular capacitance symbol with an arrow across it, or a capacitance symbol with one curved side with an arrow attached to it.



variable capacitor

Sometimes, when a very small capacitance needs to be adjusted a single time, the capacitor will take the form of a short pair of twisted wires. Twisting the wires increases the capacitance. Techs call this a **gimmick**.

gimmick

There is also a solid-state device called the **varactor** whose capacitance can be varied by changing the voltage across it. We will study it when we get to semiconductors in EM3.

CAPACITIVE REACTANCE

Like a resistor, a capacitor opposes current flow. However, it differs significantly from a resistor:

- Its opposition to current flow varies with frequency; as frequency rises, opposition falls.
- It does not dissipate power, but stores and releases it.
- AC voltage across a capacitor and current through it are not in step with each other.

Because of these differences, we don't call a capacitor's opposition to current “resistance” but **reactance**. The symbol for reactance is **X**. There are two kinds of reactance, capacitive and inductive, so we add a subscript: the symbol for capacitive reactance is $\mathbf{X_C}$. Reactance is measured in Ohms. I'll have a lot more to say about this in the next chapter, where we study capacitors in detail.

INDUCTORS

An inductor is a coil of wire, often with an iron piece in the center. The name comes from the fact that a magnet moving near it will “induce” a voltage in it, and a current through it will “induce” a magnetic field. Its electrical behavior is called **inductance**, which for some reason is abbreviated **L**. The unit of measure for inductance is the **Henry**, abbreviated **H**, as in L = 3H. Smaller units of inductance are the **millihenry**, abbreviated **mH**, and the **microhenry**, abbreviated **μH**.

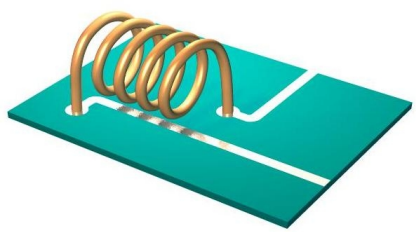
The inductor is the mirror image of a capacitor: the same in every way, yet opposite in every way. The energy in an inductor is kinetic energy. An inductor has electrical inertia. To paraphrase Sir Isaac Newton, in an inductor, electrons at rest tend to stay at rest; electrons in motion tend to stay in motion.

Inductors depend on electromagnetism. If you've still got the EM1 study guide sitting around, you may want to quickly review the magnetism chapter. In that chapter, we saw that current produces a magnetic field. Inductance comes from the interplay of changing currents and changing magnetic fields, as I shall explain in wonderfully lucid detail in chapter seven.

Inductor Construction

A straight wire has some inductance, a fact that designers of high-frequency systems continually have to wrestle with. But a component that is deliberately designed for inductance almost always takes the form of copper wire wound into a coil. You know from EM1 that each loop of wire is called a **turn**. The turns must be prevented from shorting to each other, so the wire must be coated with insulation, usually plastic.

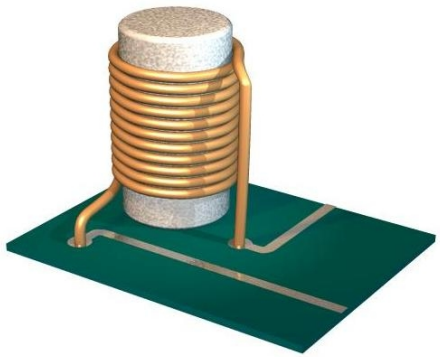
Air-Core Inductors



The space inside the coil is called the **core**. The simplest core is just air. Air-core inductors are used where we need just a little bit of inductance.

At very high RF frequencies, an inductor is often just a few self-supporting turns of sturdy wire with the ends soldered into a PC board. Such a component might have an inductance of a few μH.

A larger inductance requires more turns, which might require a physical support. In this case the coil is wound on a paper, ceramic or plastic tube. This is equivalent to an air-core inductor, since plastic, ceramic and paper are nonmagnetic and nonconducting and thus invisible to electromagnetism.



Where a lot of inductance is required, an air-core inductor won't work. So many turns of wire would be required that the wire would either have to be ridiculously thin, meaning that its resistance would be unacceptably high, or the coil would have to be huge. But a huge coil implies big turns meaning that each turn would have a comparatively long length of wire, again increasing the resistance. We could reduce the resistance by using thicker wire, but that would make the coil even bigger, requiring more wire, which would increase resistance...

In other words, you reach a point where you can't win. Everything you do defeats something else you did.

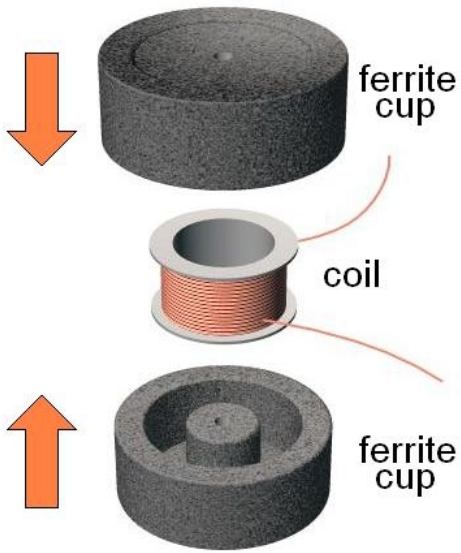
However, there is a way of obtaining much larger inductances. In EM1, we saw that inserting a soft iron core in an electromagnet makes the field stronger, and creating an entire magnetic circuit makes the field stronger still. These measures also increase inductance. In this way, we can obtain inductances of multiple Henrys. Only “soft” ferromagnetic materials can be used, because they must magnetize, demagnetize and reverse-magnetize with every AC cycle. Hard iron and its alloys, which make permanent magnets, cannot be used for inductor cores.

Ferrite Cores

Ferrite is a material made of soft iron or iron alloy dust mixed with a semi-liquid ceramic gunk that is molded into the desired shape and then hardened. Since only a part of the ferrite material is actual iron, it's not as magnetic as pure iron. An inductor with a ferrite core has more inductance than an equivalent air-core inductor, but less than one with a soft iron core. Ferrites are used at low radio frequencies, e.g. in modern power supply circuits that generate high-power square waves at 50kHz and higher, and on up through hundreds of MHz.

The illustration at right shows how a ferrite-core inductor used in a power supply is assembled from a coil and two half-cores. Note that the assembled unit constitutes a full magnetic circuit.

Ferrite materials are brittle and will shatter if struck by a hard object or physically stressed.

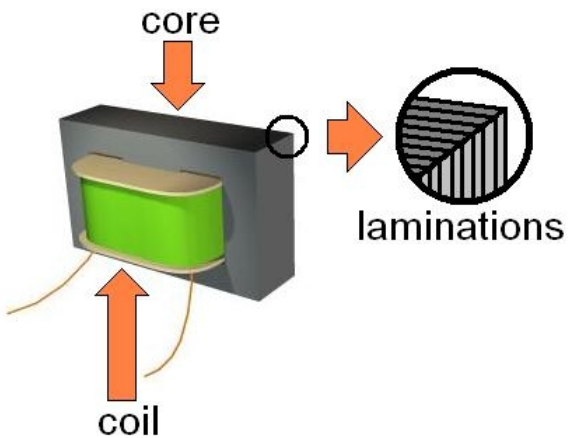


Iron Cores

The highest inductance is obtained with a core of soft iron, especially if it's formed into a magnetic circuit that surrounds the coil. Iron-core inductances range from hundreds of mH to tens of Henrys. They're used at audio frequencies, especially in 60Hz power systems, and sometimes even at infrasonic frequencies. Iron cores cannot be used above about 50kHz because internal losses become too great.

Laminations and Eddy Currents

An iron core will be divided into **laminations**, which are thin sheets coated with a a plastic film that electrically insulates them from each other. This is to prevent **eddy currents**, which are circulating currents inside the iron core caused by the alternating magnetic field. Eddy currents waste power and heat up the core, thus the need to break up the core into laminations so magnetism can flow in it but electricity can't. In some cases, instead of thin sheets, the laminations take the form of a single long sheet wrapped up in a coil. This results in a donut-shaped core, called a **toroid**.

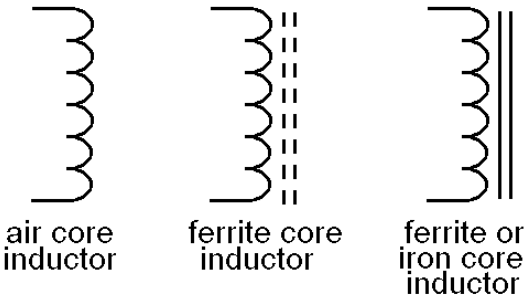


Ferrites are largely made of insulating material, so eddy currents are not an issue with ferrites.

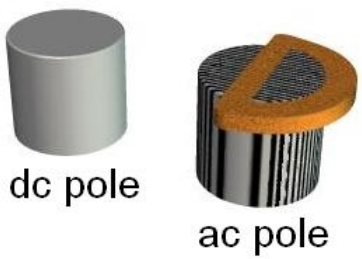
An inductor has no polarity and may be connected either way.

Schematic Symbols

The symbol for an inductor suggests a coil of wire. An iron core, if present, is represented by a pair of lines next to the inductor. If the core is ferrite, it may be represented by dashed lines. You've seen the symbol on the right in EM1 as part of the schematic symbol for a relay, which has an iron-core coil in it.

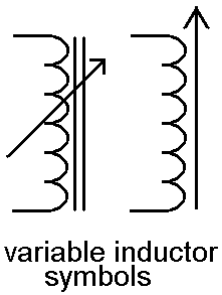


Side note – you can usually tell whether a relay coil is intended for AC or DC by the pole piece. If it's a single piece, it's a DC coil. If the pole piece is laminated, it's an AC coil. This is especially true if part of the pole piece has a copper ring around it. This is called a **shaded pole** and we'll cover it when we talk about motors in EM5. The purpose of the shaded pole is to keep the relay from buzzing.



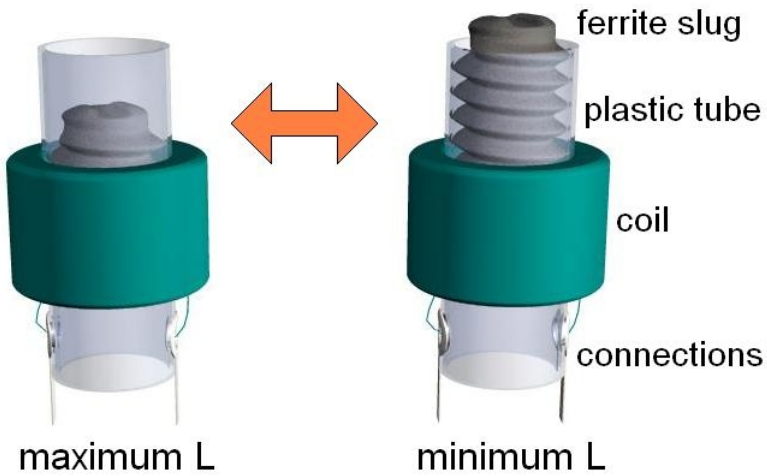
Variable Inductors

The easiest way to vary inductance is to choose an inductor with an iron or ferrite core, and move the core in and out of the coil. When the core is in the coil, it's an iron- or ferrite-core inductor with high inductance; when it's out of the coil, it's an air-core inductor with low inductance, and when it's partway in the inductor has intermediate inductance.



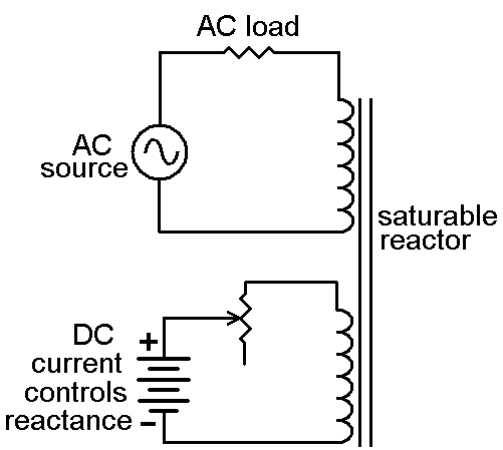
The most common type is a coil wound on a paper or plastic tube, with a ferrite core that has screw threads molded into it (see illustration next page). A slot or six-sided hole is molded into the top of the core, which permits a screwdriver or allen wrench to screw it in or out, changing the inductance. This must be a plastic tool, both because a steel tool will throw off the inductance, and because a steel tool may crack the brittle ferrite.

A ferrite core of this type is often called a **slug**. This kind of variable inductance is often used in the tuned circuits of radios and elsewhere, in which case the method of adjusting the tuning is called **slug tuning**.

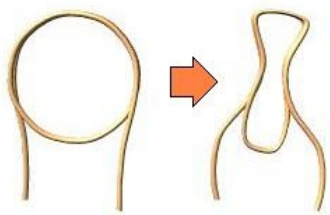


Saturable Reactors

Another way to vary inductance is to deliberately saturate the core. This is mostly used with soft iron cores, as ferrite cores are hard to saturate and air does not saturate. A second coil is wound on the core, and an adjustable DC current is passed through it. If the DC current is set to the right value, the core will begin to saturate, which reduces its inductance. More or less DC means more or less complete saturation. This is used in older lamp dimmers and machine controls.



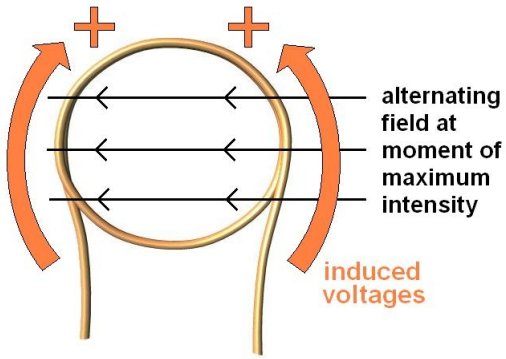
If we're just making a one-time adjustment of a self-supporting coil, the adjustment may be as simple as stretching the coil or twisting it out of shape and then dripping a glob of wax or hot glue on it so it will stay, and so vibration will not cause momentary distortions in the coil, which would cause momentary changes in inductance. (Electrical disturbance caused by mechanical vibration is called **microphonics**.) This kind of one-time adjustment is done a lot in non-digital FM and TV tuners. Pinching a loop of wire so it's no longer a wide loop also reduces the inductance. Don't tell anybody I showed you that.



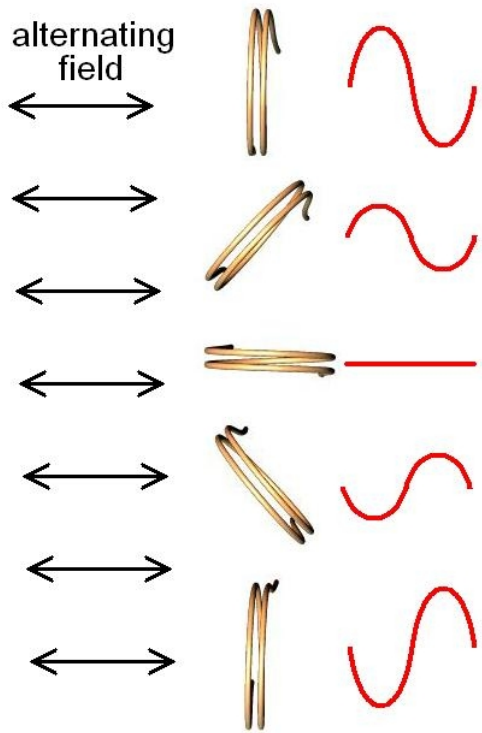
INDUCTIVE PICKUP FROM AN EXTERNAL FIELD

You know from EM1 that an alternating magnetic field will induce a voltage in a coil. We've seen that the field that a coil creates is oriented on its axis, so the lines of force go through the loop. This is also important going the other way, when an external field is inducing a voltage in a coil.

Inductive pickup only works when lines of force travel *through* the center of a coil. If they travel *across* a coil, as in the illustration at right, nothing happens. It is true that, even when the field travels across the coil, it's traveling across large parts of the wire, and voltages are induced in those sections. But the voltages induced in one side of the coil are opposite in polarity from those in the other side, so they cancel.



In fact, if you rotate the coil and compare the phase of the voltage to the field, as in the illustration at right, you'll find that it goes from maximum voltage to zero to maximum again, but opposite polarity. Voltage is proportional to the sine of the angle; it's greatest when the coil is at 90 degrees to the lines of force. Zero degrees, the point where there's no voltage, is called a **null**. A null occurs when two things – two fields, two voltages, whatever – cancel each other out.



INDUCTIVE REACTANCE

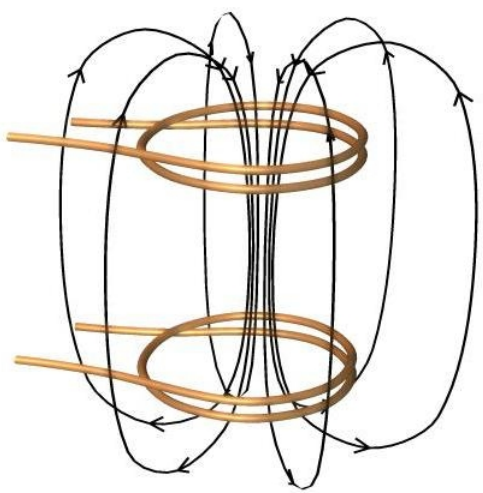
An inductor, like a capacitor, has reactance, not resistance. This means:

- Its opposition to current flow varies with frequency, but in the opposite direction from capacitors: as frequency rises, opposition rises.
- It does not dissipate power, but stores and releases it.
- AC voltage across an inductor and current through it are not in step with each other.

To distinguish inductive reactance from capacitive reactance, we use a subscript: $\mathbf{X_L}$. Inductors and their properties will be examined in greater detail in chapter 7.

MUTUAL INDUCTANCE

When two inductors are close enough together that their fields overlap, the part of the field they share gives rise to **mutual inductance**. Depending on how they are connected, the inductance of each may increase the inductance of the other. The part of the inductance that is not shared is called **self-inductance**.



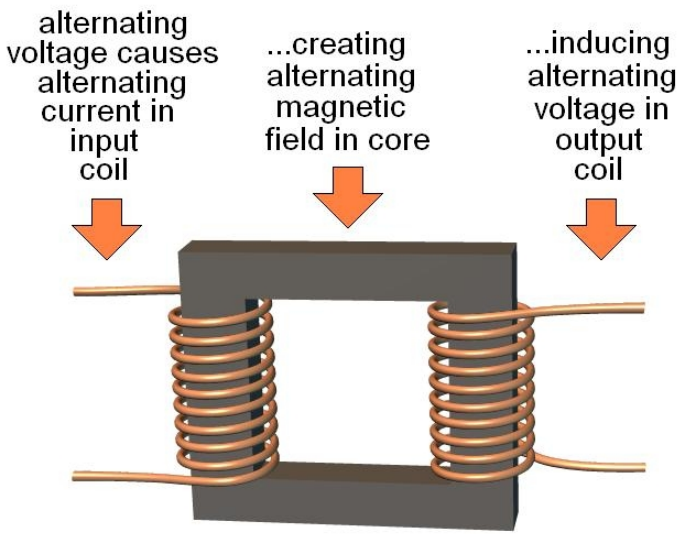
ENERGY TRANSFER VIA MUTUAL INDUCTANCE

An alternating current through a coil creates an alternating field. If this coil shares part of its field with a second coil, that means the second coil will be in an alternating field too, and a voltage will be induced in that coil. It is not hard to see that when two coils share a field, energy can be transferred from one to the other in this manner. This occurs without any electrical connection between the coils; the energy transfer is accomplished magnetically. There is a double transformation: electricity in one coil to magnetism, then back to electricity in the other coil. The device that is built to make maximum use of this effect is called a **transformer**.

TRANSFORMERS

Long before “Transformers” was a dumb cartoon and a series of brain-dead movies – I don't care if you liked them, I'm entitled to my opinion – it was a name given to a type of energy transfer device.

A transformer consists of two coils of wire so arranged as to maximize their mutual inductance, ideally to the point where the individual coils don't have independent fields at all. When an AC voltage is applied to one coil, it creates an alternating magnetic field. This field induces an AC voltage in the second coil. Telephone engineers call this a **repeat coil**, which seems like a pretty good description.

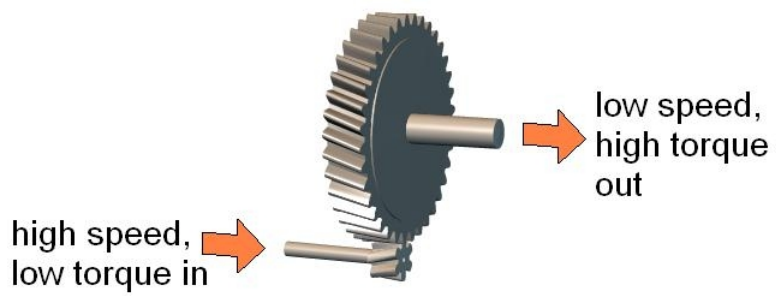

alternating voltage causes alternating current in input coil
...creating alternating magnetic field in core
...inducing alternating voltage in output coil

transformer operating principle

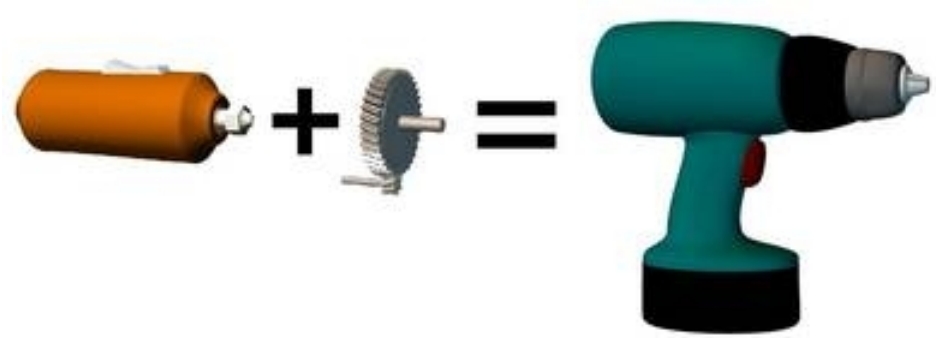
So, you put in an AC voltage and you get the same AC voltage out again. That's gotta be as useless as a whoopie cushion at a wake, right? “Hey maw, we're gonna be rich. I got me a dee-vice where, you put in a volt, you get out a volt. Everybody'll hafta have one o' these!”

Well, actually, it's very useful. First off, note that a transformer can get electricity from one place to another without any actual electrical connection between the source and the load. When the source is a 117V outlet and the load is a metal-cased appliance you're holding in your hand, that isolation is a nice safety feature. Computers, music systems, guitar amps, TVs, test equipment etc have transformers in their power circuits partly for this reason.

Second, it's possible to make the transformer so a different voltage comes out than what went in. In this way a transformer is like a gearbox. With a rotating gizmo, power is speed in RPM times torque in foot-pounds or whatever. A gearbox does not create power, and if we ignore friction, does not lose it. The power that goes in is the power that comes out. What a gearbox does is change the speed-to-torque ratio.



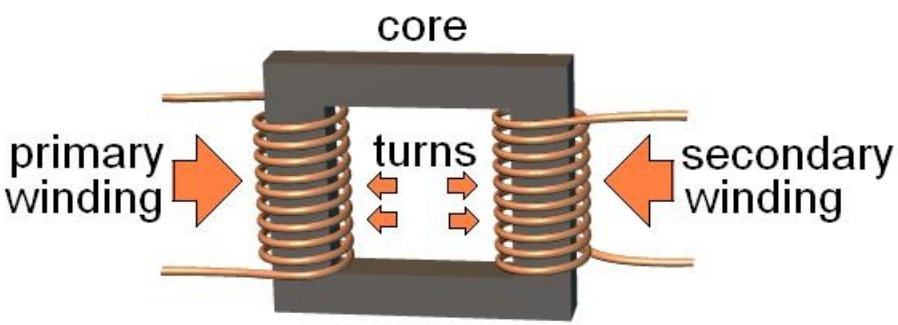
Consider the Dremel tool. Okay, to keep the lawyers off my back: consider a Dremel-brand high-speed rotary tool. It spins at up to 25,000 RPM but it doesn't have a lot of torque. Now consider a hand drill. It spins much more slowly, maybe 1,000 RPM, but it has much greater torque.



But the two tools have essentially the same motor and produce about the same power. The difference is, there's a speed reducing gearbox in the hand drill that trades speed for torque. The ratio of input to output speed is the “gear ratio” and in this case it would be 25-to-1.

The electrical equivalent to that gearbox is the transformer that knocks the 117VAC from a power outlet down to 12V to run the halogen light in your desk lamp. It trades high voltage at low current for lower voltage at higher current; ideally, ten watts from the outlet, ten watts into the lamp. (Of course, there is always some loss in the real world. In a transformer, this is called **insertion loss**.)

Terms: in a transformer, each coil is called a **winding**, pronounced like “finding.” The input coil is called the **primary winding** and the output coil is called the **secondary winding**. Each loop of wire in a winding is called a **turn**. If there's a piece of ferromagnetic material involved, it's a **core** just as with an inductor. Transformers that work at very high frequencies may have an air core. If the output voltage is higher than the input voltage, it's a **step-up** transformer. If the output voltage is lower than the input, it's a **step-down**. Just as with a gearbox, the numbers are stated as a ratio.



TOLERANCES

As you'll recall from EM1, the specified values of parts are not exact. A part will have a specified **tolerance** which tells you how far from the stated value the actual value can be.

Dry capacitor tolerances are 5% or wider. Large disc caps, say 0.01μfd and larger, are quite sloppy, display considerable drift with change in temperature, and can distort waveforms. They should not be used except where crude, bulk capacitance is acceptable. Capacitors using mica, glass and poly plastic dielectrics are much better. Electrolytic caps can be off as much as 50% (no lie!). That doesn't mean they're low quality; it's just the way electrolytics are. They also change with age, though **tantalum** electrolytics are better in that respect. Electrolytics simply shouldn't be used where their exact value makes a difference.

Inductor tolerances are usually around 5%.

Higher precision resistors, capacitors and inductors are available at exponentially higher prices. Alternately, small variable components can be used to trim larger components to precise values, as described above and in EM1.

Transformer step-up or step-down ratios are usually within 5% of specification, but the specification is taken with the transformer fully loaded. With a power or audio transformer, a lighter load will result in a smaller insertion loss and thus a slightly higher output voltage. With RF transformers, you don't usually have to consider what will happen if the source and load impedance vary, because they don't. For reasons discussed in chapter 10, RF transformers are almost always used in circuits where source and load impedances are tightly controlled.

The next three chapters take capacitors, inductors and transformers respectively and look at them in more detail.

End of Chapter 5

EM2 Chapter 6
PRINCIPLES OF CAPACITANCE

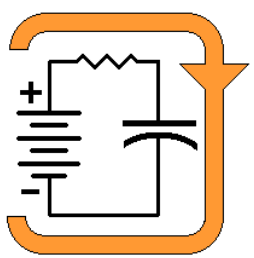
Objectives of this chapter:

- *describe the proportion and phase relationships among E, I and R in a capacitor*
- *solve series and parallel capacitance problems*
- *discuss "reactance" and "impedance"*
- *describe "time constant" and "charging time"*
- *discuss blocking, coupling, decoupling and filtering uses*

CAPACITOR OPERATION

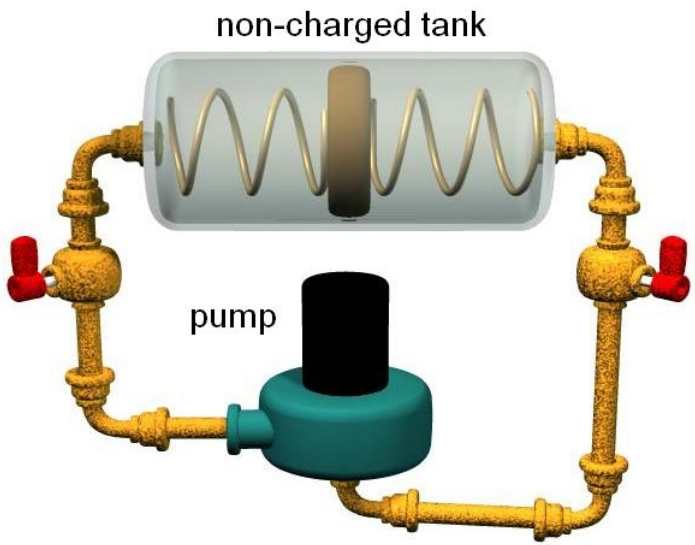
When current flows through a capacitor, it charges or discharges – that is, the potential across it increases or decreases.

How could current flow *through* a capacitor? After all, it consists of two plates with no electrical connection between them; it's an open circuit. The answer is, current can't flow continuously. But during the time when a capacitor is charging or discharging – during the time when the voltage across it is changing – electrons are flowing into one plate and out of the other.

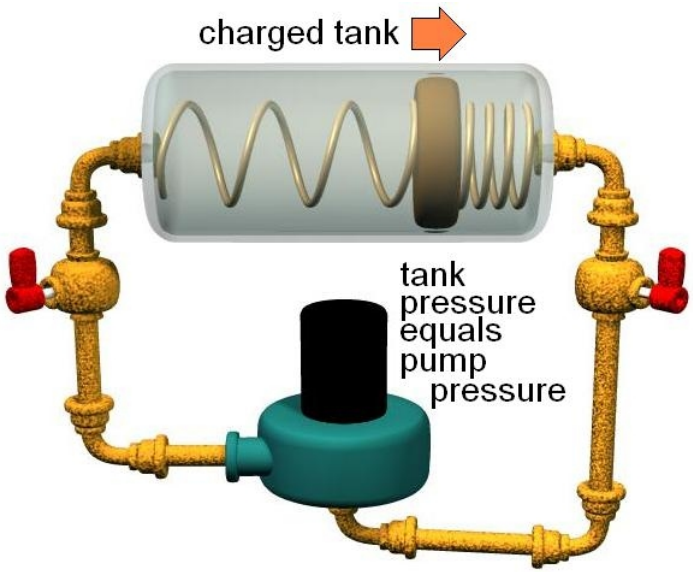


current flows thru cap while charging or discharging

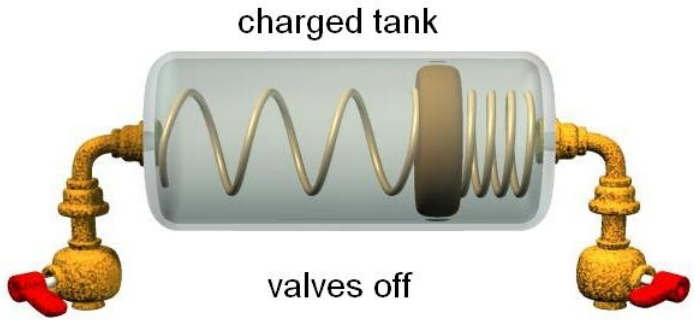
Think of a capacitor as a water tank with a pipe attached to each end. In the middle of the tank, there's a plug blocking water flow, so water can't flow continuously. (That's the open circuit.) Water pressure can push the plug back and forth, but can't get past the plug. Under normal conditions, the plug is kept centered by springs.



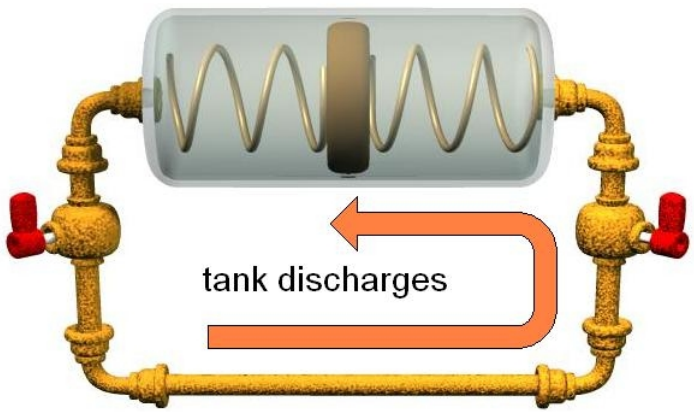
The tank is connected to a pump, with valves between. When the pump turns on, water will flow through the tank for a while, which will push the plug to one side. When the plug is pushed far enough that the spring tension balances the pump pressure, the plug comes to rest. The spring tension, pushing on the plug which pushes on the water, is equivalent to the charge on a capacitor.



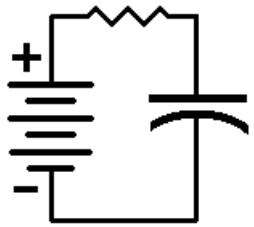
Let's close the valves, so water can't flow. If we remove the pump, the pressure differential inside the tank – the charge – remains.



Now let's connect a length of pipe where the pump was and open the valves. The springs push the plug, which pushes the water, which flows through the pipe, allowing the plug to return to its original position.



Let's look at an electrical circuit that works the same way. The pump is replaced by a battery, the tank by a capacitor, and constriction of the water flow – there's always some – is represented by a resistor. We can state these facts:



- in a capacitor, energy is stored as a charge difference across conductive plates.
- as current flows through a capacitor, voltage across it changes; the capacitor **charges** or **discharges**.
- the larger the current, the faster voltage changes.
- the larger the capacitor, the slower voltage changes. Think of it as a larger water tank that takes more water to fill or empty.
- the greater the charge on the capacitor, the higher the voltage across it.
- when the voltage across the capacitor equals the battery voltage, charging stops, current flow stops, and the capacitor voltage stabilizes.
- if the battery is disconnected from the circuit, there's no way for current to flow. Whatever charge is on the capacitor at that time will stay there. (We're talking about ideal capacitors. Real capacitors have a bit of leakage, so they slowly discharge by themselves.)
- if a current path is connected across a charged capacitor, the capacitor will push current in the other direction and will discharge.

Time to get more specific. While current through a resistor is proportional to the voltage across it, current through a capacitor is proportional to the *rate of change* of voltage. Stated as a formula:

$$\Delta V/\Delta T = I/C$$

where V = potential in volts, T = time in seconds, I = current in amps, and C = capacitance in farads. The triangles are Greek Deltas. In math and electronics, a delta signifies change.

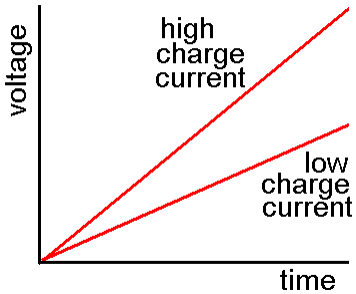
In plain English, the change in voltage across a capacitor per second equals charging or discharging current divided by capacitance. This formula demonstrates two relationships already mentioned: the greater the current, the faster the voltage change; the greater the capacitance, the slower the voltage change.

The equation is also commonly written:

$$dV/dT = I/C$$

where “d” means “difference.”

If you charge a capacitor with a constant, unchanging current, the voltage across it rises at a constant, unchanging rate. The higher the current, the faster the voltage will rise. And, if you keep it up, the voltage will rise until the dielectric breakdown voltage is reached and the capacitor is destroyed. There is no built-in leveling off point; it's a simple linear function.

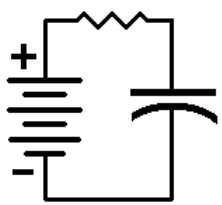


This means we can find the voltage across the capacitor at any moment by rearranging the previous equation. Assuming we start with zero volts across the capacitor:

$$V = I/C \times T$$

where I is current in amps, C is capacitance in Farads, and T is time in seconds.

How do you get a constant current? You study EM3, where we'll discuss constant-current circuits. (They require transistors etc.) Much more often, capacitors are charged through a resistor. In this case, the charge rate is not constant. The resistor and capacitor form a kind of voltage divider. As the capacitor charges, the voltage across it increases, thus the voltage across the resistor decreases, thus the current decreases, thus the charge rate gets slower and slower. In theory, the capacitor never does fully charge up to the same voltage as the source, just as, if you pour enough water into a glass to make it ½ full, then $^2/_3$, then ¾, then $^4/_5$ and so on, you'll never fill it. So we have to talk in terms of how long it takes a capacitor to become "mostly" charged.



RESISTANCE-CAPACITANCE TIME CONSTANTS

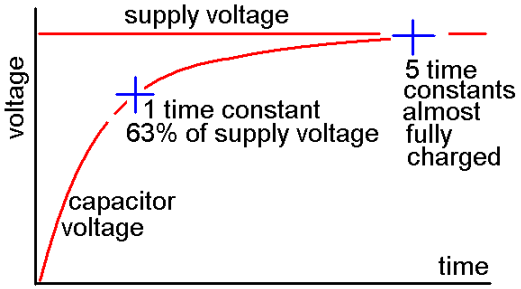
When a capacitor and a resistor are connected together, the combination has a characteristic **time constant**. Mathematically, this is one of the simplest formulas you'll ever run into:

$$T = RC$$

Time in seconds, resistance in ohms, capacitance in farads. The time constant is the time it takes for a capacitor to reach 63% of the supply voltage.

Why 63%? Because it works out in equations, for reasons I'm not going to go into right now. But 63% is close to $^2/_3$, so you can think of the time constant as the time it takes for $^2/_3$ of the charging process to happen.

The chart at right shows how voltage varies with time. Charging rate is fastest at the start, when current is greatest, and then slows down. At one time constant, voltage across the capacitor has reached 63% of the supply voltage.



Since the capacitor never reaches the full supply voltage, we have to make an arbitrary decision about when we're going to call it fully charged. In practice, a capacitor is considered "as good as charged" in five time constants, according to the equation:

$$T_{FULLCHARGE} = 5\,RC$$

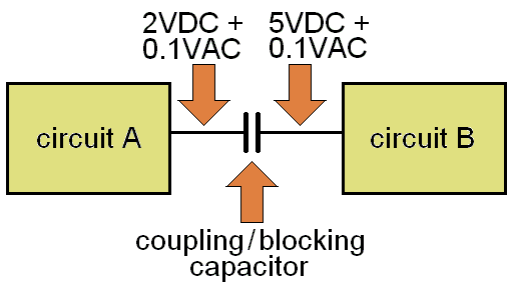
where T = time to essentially full charge, R is resistance in Ohms, and C is capacitance in Farads.

INSTANTANEOUS DISCHARGE

What happens if you put a screwdriver across a capacitor, discharging it instantly? Wouldn't the instantaneous voltage change correspond to an infinite current? Yep. There's no such thing as infinite current in the real world, of course; a screwdriver has a low but significant resistance and the capacitor itself has some. But a lot of technicians own pitted screwdrivers proving that big capacitors have very low internal resistance and can discharge at an impressive rate of current. This is even used as a spot-welding method.

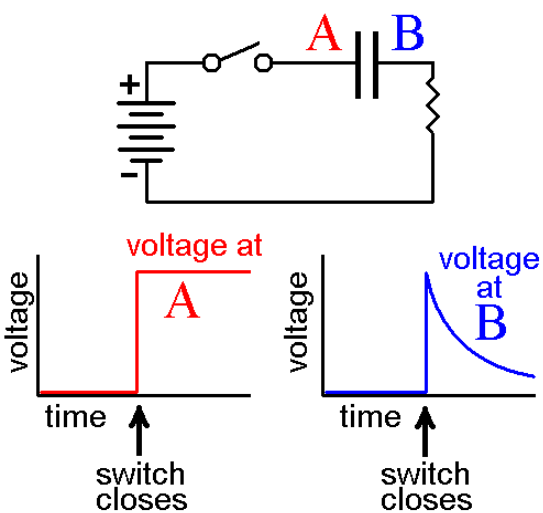
CAPACITORS AND ALTERNATING CURRENT

Capacitors have uses besides storing a charge. One of the most common is to couple an AC signal between two parts of a circuit. Coupling wouldn't seem to be a problem; we should be able to just connect things together. But different parts of a circuit often cannot be directly connected because their DC voltages are different, as in the illustration where we want to connect two points that have, respectively, 2V and 5V DC on them. Capacitors to the rescue: a capacitor can *pass* or *couple* an AC signal, and at the same time, *block* a DC voltage. A capacitor used this way is called either a **coupling** capacitor or a **blocking** capacitor, as it couples the AC signal but keeps the DC potentials separate.

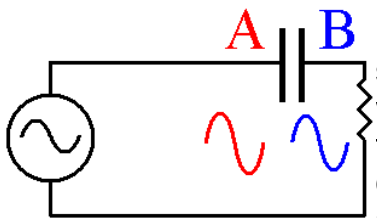


Let's see how this happens. For a moment, we're going to return to electron flow, negative to positive. Fasten your snaps, this is going to be a slightly bumpy ride.

Suppose we have a discharged capacitor connected in a circuit as shown. If we close the switch, electrons will flow out of the left-hand plate toward the positive end of the battery. The plate will become positively charged. This will create a positive field, which will attract electrons in the right-hand plate, pulling them toward the left-hand plate and putting a positive voltage at the top of the resistor. In other words, when the voltage on the left plate changes, the change will be coupled across the capacitor to the right plate and anything connected to it. Points A and B will go in the positive direction together. For the tiniest instant, the capacitor behaves almost like a wire.



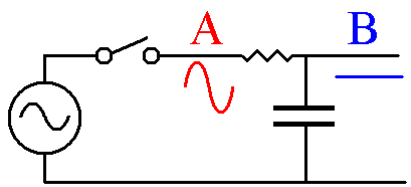
But the positive voltage on the right plate won't last. As soon as there is a positive voltage across the resistor, electrons will be drawn up through it. They will flow to the right-hand plate, and after a while, they will neutralize its positive charge, so the voltage at point B will be zero. (What's “a while?” Five RC time constants.) At that point, the capacitor is charged, with 9V at A and 0V at B. Since there is no longer any voltage across the resistor, current will stop flowing. So, over the long term, the capacitor behaves like what it is: an open circuit.



Now let's remove the battery and switch and substitute an AC voltage source. If the AC frequency is high enough, the charge on the left plate will alternate so rapidly that the transferred charge on the right plate won't have time to drain off through the resistor. So, the right-plate voltage will almost perfectly follow the left-plate voltage. Thus we have a device that will pass an AC signal but block a DC voltage; one that is like a wire at high frequencies and an open circuit at DC.

Does this AC coupling behavior mean electrons really flow through the capacitor? No. Some electrons go into storage on one side and other electrons come out of storage on the other side. But one electron is the same as another, so for all practical purposes, you can regard AC current as going through the capacitor.

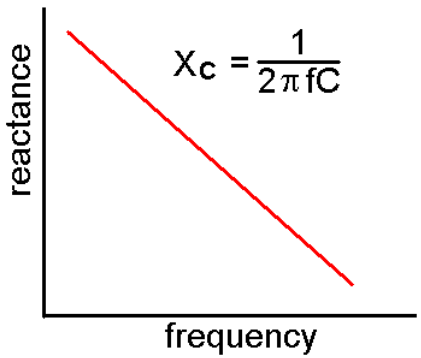
A capacitor can also be used the opposite way, to pass DC and confine AC, by connecting it across the voltage. This only works when there is significant resistance in the circuit. At high frequencies, the capacitor will short out alternating currents, but have no effect on DC as the capacitor is an open circuit at DC. One way this is used is to keep AC voltages from accidentally being coupled between sections of a piece of equipment via the power supply rails. A capacitor used this way is called a **decoupling** or **bypass** capacitor.



Capacitive Reactance

Let's take what we've covered so far and broaden it a bit. A capacitor, like a resistor, presents a resistance to current flow. But “resistance” is the wrong word. A capacitor does not, ideally, dissipate energy as a resistor does; with AC, it alternately stores and releases energy. So, the kind of “resistance” a capacitor presents has been given a different name: **reactance**. Reactance is enough like resistance that it can be measured in Ohms. The symbol for reactance can't be R, because that's already been used for resistance, so “**X**” has been chosen. (Of course this means we can no longer use “X” as a general variable in equations. It's also a good reason to think twice about putting an “x” into a formula to indicate multiplication.) We have to be a bit more specific, because there are two kinds of reactance, so we use a subscript. **Xc** means **capacitive reactance**.

What makes reactance different from resistance, aside from the fact that it doesn't dissipate power, is that it changes with frequency. With capacitive reactance, it's an inverse relationship: the higher the frequency, the lower the reactance. If the frequency is high enough, reactance is almost zero and the capacitor is almost like a piece of wire. At DC, which can be thought of as the lowest possible frequency, reactance is infinite; that is, an open circuit.



Also, the larger the capacitance, the smaller the reactance at any particular frequency, another inverse relationship. We can put these relationships into a formula:

$$Xc = 1 / (2\pi fC)$$

Where Xc is capacitive reactance in Ohms, f is frequency in Hz, and C is capacitance in Farads.

In this form, the equation is nearly useless, because the capacitors used in AC applications are seldom anywhere near a Farad. Instead, we use these formulas:

$$Xc = 10^6 / 2\pi fC$$, where C is in microfarads, and

$$Xc = 10^{12} / 2\pi fC$$, where C is in picofarads.

If you mastered the EM1 math chapter you'll remember that a micro-anything is $10^{-6,}$ a pico-anything is 10^{-12}, and in scientific notation a negative power of 10 in the denominator can be exchanged for a positive power of 10 in the numerator if the denominator contains only multiplications, not additions.

Observe that both of the variables in the fraction are below the line, in the denominator. That's because capacitive reactance is *inversely* proportional to frequency and capacitance. This is a potential point of confusion, because it's "upside down" from resistance. A larger resistor value means more resistance, but a larger capacitor value means *less* reactance.

Where'd that 2π business come from? Well, recall from the math chapter that AC sinewaves correspond to the circular motion of phasors, that a phasor is the radius of a circle, and the length of a trip around the circle – one cycle of a sinewave – is 2π times the length of the radius, or 2π radians. What we're doing is adding a conversion factor to the equation so instead of talking about angular velocity, we can talk about frequency, which is more useful.

Let's work an example. Suppose we want to find X_C of a 1μF capacitor at 1kHz. We start with this equation:

$$Xc = 10^6 / 2\pi fC$$

and plug in the values we know, then proceed on through the steps:

$$Xc = 10^6 / 6.3 \times 10^3 \times 1$$
$$Xc = 10^3 / 6.3$$
$$Xc = 160$$ (pretty close)

So, at 1kHz, a 1μF capacitor has 160Ω of capacitive reactance.

Just to lock these concepts in, particularly the "upside down" nature of capacitance, let's scale these quantities up and down a bit.

The size of a capacitor is inversely proportional to its reactance; at 1kHz, a 2μfd capacitor has only 80Ω of reactance, while a ½μfd capacitor has 320Ω.

The reactance of a capacitor is inversely proportional to the frequency; at 2kHz, a 1μfd capacitor has only 80Ω of reactance; while at 500Hz it has 320Ω.

Instead of calculating, you may find it easier to memorize a particular capacitor's reactance at a particular frequency, and then **scale** up or down. Specifically:

0.016μfd, 10kΩ, 1kHz

A 0.016μfd capacitor has 10kΩ of reactance at 1kHz. You can scale from there – if frequency or capacitance goes up, reactance goes down, and vice versa. This is especially handy when you have to calculate in your head. I use the scaling method instead of the equation most of the time.

CAPACITORS IN SERIES AND PARALLEL

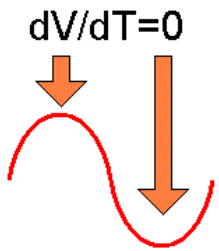
Capacitive reactances in series and parallel work like resistances in series and parallel. When capacitors are connected in series, total X_C is the sum of individual X_C's, and is always higher than any individual X_C. When capacitors are connected in parallel, the total X_C is always lower than any individual X_C, and you find it using the same product-over-sum or reciprocal formula used with resistances. The difference is that you have to keep in mind that, like an individual X_C, the number will change with frequency.

VOLTAGE AND CURRENT PHASE

The AC voltage across a capacitor and the current through it are not in phase. If the voltage is a sine wave, current will also be a sine wave, but it will lead the voltage by 90 degrees. I mentioned this in chapter 5, and now that you know more about capacitors, I can show you why this happens.

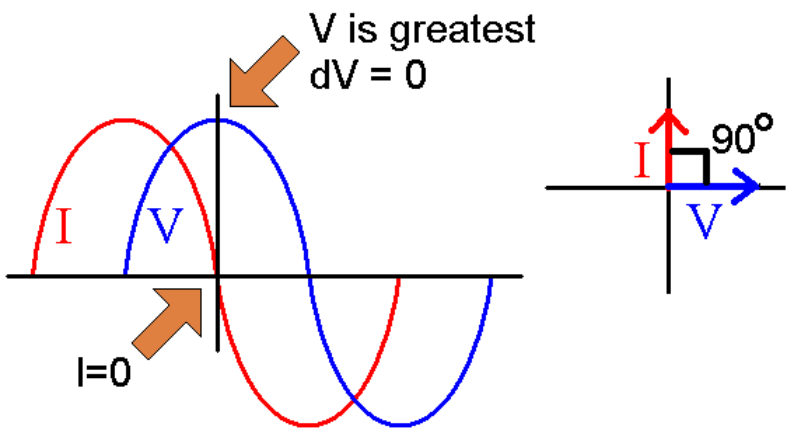
You know that the rate of change of voltage across a capacitor, dV/dT, is proportional to current. Faster voltage change, higher current, as the capacitor charges or discharges; slower voltage change, lower current.

Where, in a sine wave, is the rate of voltage change smallest? Right at the positive and negative peaks. In fact, there's a moment at each peak when the rate of change, dV/dT, is zero, as the voltage is in the process of changing direction.



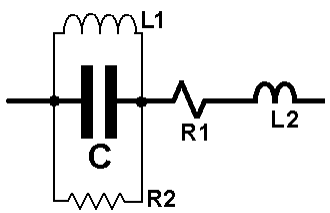
So, let's connect a sine wave voltage across a capacitor. What would the current through a capacitor be at the voltage peaks, when voltage isn't changing? Well, when voltage across a capacitor isn't changing that means no electrons are flowing in or out, so current is zero.

If we draw a current sine wave on top of the voltage sine wave, we will see that it is displaced in time so that the moments when current is zero coincide with the moments when the voltage is at its peaks. In other words, there will be a 90-degree difference, with current leading voltage.

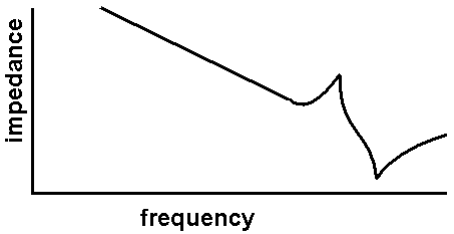


REAL-WORLD CAPACITORS

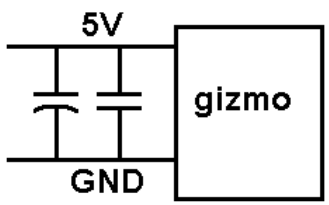
Actual capacitors aren't pure capacitance. They have resistances that cause losses, and inductances that upset the clean X_C vs. frequency relationship. These are called **parasitic** resistances and inductances.



We haven't dealt with resonance yet – that comes up in chapter 9 – but the parallel circuit of C and L_1 will, at some high frequency, cause the capacitor's impedance to rise. On the other hand, the series circuit of C and L_2 will, at some other frequency, cause the impedance to dip, but will limit how low the impedance will go at still higher frequencies. This results in an impedance curve that, at high frequencies, resembles the one at right.



One of the results of this is that you'll often see two capacitors, an electrolytic and a dry capacitor, across the power supply next to an IC. These are bypass capacitors, and their purpose is to make sure the power supply is very stiff at that point, so varying loads do not result in varying power supply voltage.



The electrolytic, being a comparatively big capacitor, ought to bypass everything nicely. But just by virtue of those rolls of foil inside, it has enough inductance not to make a very good bypass at very high frequencies. So it's augmented by a small dry capacitor. You'll see this next to microprocessors in particular.

DIELECTRIC CHARGING

I know I'm going to be told that the following is esoteric and doesn't belong in an introductory study guide, but I'm putting it in because it has to do with safety.

I mentioned that charge isn't only stored on a capacitor's plates, the dielectric stores it too. This occurs because the dielectric lives in the electrical field between the plates, and the orbiting electrons in its atoms get pushed a little bit off center.

However, if the charge lasts long enough, the dielectric becomes physically charged. No insulator is perfect. Electrons slowly creep through the dielectric and give it a static charge, and since we're talking about a charge difference – a potential – the charge can be very significant. The late analog guru Bob Pease called this effect "soakage," because electrons slowly soak through the dielectric like water through a sponge.

When the capacitor is discharged, the opposite effect occurs. Suppose we short-circuit the plates – the immediate charge on the plates is drained off, but the charge in the dielectric remains. And, if the plates remain shorted, the dielectric charge will slowly soak its way back to the plates and be discharged.

But what if the plates do not remain shorted? The electrons soaking back through the dielectric to the plates will charge them up again.

You can verify this yourself. Connect a DC voltage across a big electrolytic capacitor to charge it up. Leave it that way for a while. Then disconnect the capacitor and short the terminals to discharge it. Remove the short, connect a voltmeter and watch what happens: the voltage slowly rises as the capacitor recharges itself.

I can think of two obvious implications. First, you learned in the safety chapter of EM1 to discharge power supply capacitors before working on a piece of equipment. If it's been a while since you discharged them, better discharge them again – they may have charged up again.

Second, be extremely careful around high voltage capacitors. Bob Pease cited a case where some big high-volt caps were disconnected and discharged, shorting bars were attached, and they were put in boxes to be moved. During the move, one of the boxes fell open and a shorting bar was knocked off. One of the movers touched the terminals and was killed. It had self-charged to that degree, even though the shorting bar had been on for hours.

Capacitors can sneak up on ya, and they don't just shout "Boo!"

End of EM2 Chapter 6

EM2 Chapter 7
PRINCIPLES OF INDUCTANCE

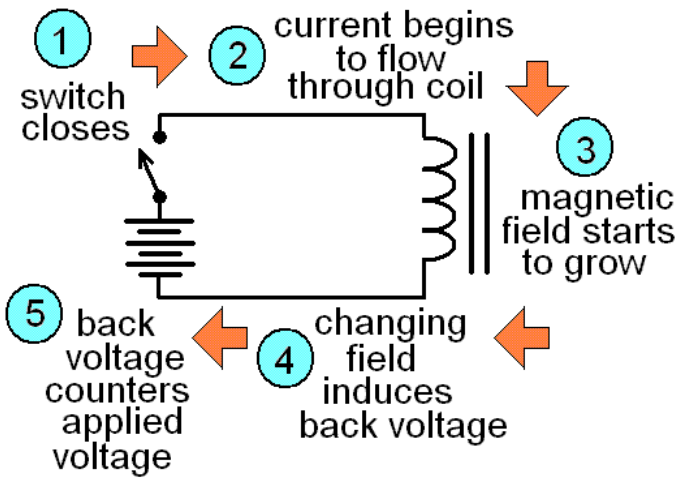
Objectives of this chapter

- *define inductance and reactance and state the relationship between them*
- *explain how the interaction of electric current and the resulting magnetic field gives rise to inductance*
- *state the formulas relating voltage, inductance, current and time*
- *explain RL time constants*
- *solve series and parallel inductor problems*
- *describe "self inductance" and "mutual inductance"*
- *describe how an inductor reacts to an AC voltage*
- *discuss Lenz' and Faraday's laws of inductance*
- *list the physical factors that determine inductance in single-layer and multiple-layer air wound coils and give the formula*
- *explain the effect of shorted turns in a winding*
- *explain the causes of power losses and how to allow for power consumption within an inductor*
- *discuss "saturation" and its effects*

INDUCTOR OPERATION

We've already noted that, whenever electrons move, they create a magnetic field. If we take a wire through which current is flowing and coil it up, this gathers and reinforces the field. When we're dealing with a DC current, a coil of wire is just an electromagnet. But with AC, a coil becomes an **inductor.** To summarize from chapter 5, **inductance** is denoted by the letter **L**, and measured in **Henrys**. An inductor has **reactance**, abbreviated $\mathbf{X_L}$, which is like electrical inertia. It is measured in **Ohms.**

We know that an alternating current produces an alternating magnetic field. We also know that a coil in an alternating field produces an alternating voltage. So, if you connect a voltage across the ends of the wire, wouldn't the changing magnetic field that results have the same effect as an external changing field? Wouldn't a voltage be generated that's mirror image of the original voltage? It would indeed. The act of connecting a voltage across an inductor causes the inductor to generate a voltage of its own. That voltage is called **back EMF** or **back voltage**, and it counters or cancels the applied voltage. This cancellation means that, in the very first micro-instant after voltage is applied, current flow is zero.

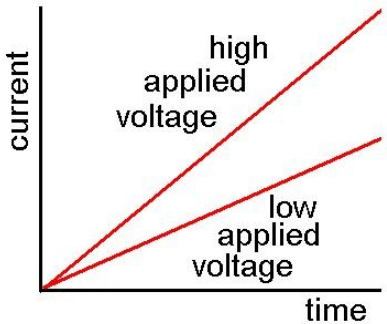


Think of two identical batteries in parallel. Since they're the same voltage, no current flows. This is the situation immediately after voltage is applied to an inductor. As soon as the voltage is applied, the magnetic field starts to build, which throws up a back voltage that counters the applied voltage. So, current = 0.

But this can't last. After all, creation of a continuous back voltage requires a continuously *changing* magnetic field. The current must increase, and keep increasing, to keep generating the back voltage.

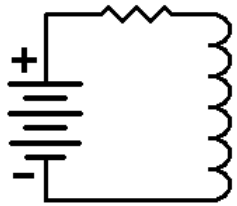
If we keep a constant, unchanging voltage connected across the coil, current will ramp up at a constant, unchanging rate to infinity, or in the real world, until the wire melts or current hits the limit imposed by wire and source resistance.

In plain English, current through a coil is proportional to the voltage connected across it and how long it's been connected, and inversely proportional to inductance. The higher the voltage, or the longer the time, the greater the current. On the other hand, the larger the inductance, the slower the rise in current. A larger inductance means that it takes less current change per time, dI/dT, to generate the necessary back voltage.



Where a capacitor is like a spring, an inductor is like a mass. It has inertia, like a car. When you step on the gas, the car's speed is zero at first and gradually increases. Likewise, when a voltage is applied across an inductor, the current is zero at first, but increases linearly with time. In other words, it accelerates.

Let's look at an electrical circuit with a DC source, a resistor and an inductor.



We can state these facts:

- in an inductor, energy is stored in the magnetic field surrounding the coil.
- whenever there is a voltage across an inductor, current changes.
- the higher the voltage, the faster the current changes.
- the larger the inductor, the slower the current changes. Think of it as a larger mass that takes more effort to accelerate or decelerate.
- when current reaches the level that the voltage drop across the resistor is equal to the source voltage, leaving no voltage drop across the inductor, current stabilizes.
- if the inductor leads are then shorted, causing voltage across the inductor to drop to zero, current will theoretically continue to circle forever, since there is no voltage across the coil and thus no change in current over time. This is not true of real-world inductors, as wire resistance will dissipate the energy. However, a circulating current can be maintained for a significant period.

- If a reverse voltage is connected across the inductor, current will ramp down in the same way that it ramped up. If a resistance is connected across the inductor, the back voltage will appear across the resistor and current will flow through it, which will dissipate the energy in the magnetic field, causing voltage to gradually (which may mean microseconds) fall to zero.

INDUCTANCE FORMULA

Voltage, current and time are related by a formula. It will be useful to compare it to the formulas we already know.

For resistance, voltage causes current:

$$V = I \times R$$

For capacitance, current causes a change in voltage:

$$\Delta V / \Delta T = I/C \quad \text{or}$$
$$V = I \times T / C$$

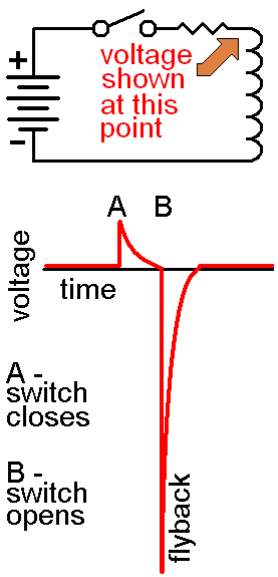
For inductance, voltage causes a change in current:

$$\Delta I / \Delta T = V/L \quad \text{or}$$
$$I = V \times T / L$$

in all cases, current in amps, time in seconds. Remember to keep the units straight: C is capacitance, not current; I is current, not inductance.

FLYBACK

We've seen that, if a resistor is connected across an inductor with current flowing in it, the current will develop a reverse voltage across the resistor, supplied by the collapsing magnetic field. What if we simply open the circuit? The current must instantly change from whatever it had been to zero.

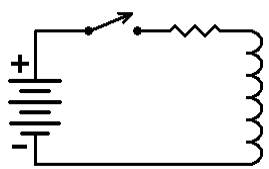


This implies that, for an instant, there must be an infinitely large reverse voltage across the inductor; theory would say this is needed to hammer the current down to zero so quickly. Where would this voltage come from? The inductor itself makes it, as the magnetic field collapses. This is called a **flyback**, and it occurs whenever a current-carrying inductive circuit is opened. The flyback as shown below left is very conservative; an actual flyback spike can be thousands of times higher than any other voltage in the circuit.

This can cause lots of problems. At the very least, it generates a burst of electromagnetic noise, which can glitch up audio, video and data. It may cause arcing. It may give somebody a shock. It may blow out equipment. Flyback suppression is one of the things engineers must do in the normal course of equipment design.

But inductive flyback can also be useful, as it gives us a way to get a burst of very high voltage from a low-voltage source. Cars use it to turn 12 volts into the thousands of volts needed to fire spark plugs. Furnaces use it to ignite fuel. Photoflash units use it to raise battery voltage to the 300-or-so volts needed to power a Xenon flashtube.

TIME CONSTANT

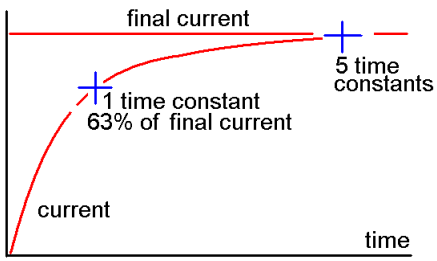


In a circuit with a resistor between the voltage source and the inductor, we get an RL time constant that's calculated almost the same way an RC time constant is:

$$T = L/R$$

where T is the time required for current to reach 63% of its final value, R = resistance in ohms, and L = inductance in Henrys. However, it operates in the opposite manner from a capacitor. Initially, when the switch is closed, no current flows, so all the voltage drop appears across the inductor and none across the resistor.

As the current through the inductor rises, the voltage drop across the resistor increases, thus the voltage drop across the inductor falls, thus the rate of current increase slows. In one time constant, current reaches 63% of its final value. In five time constants, it reaches an essentially stable level.



Just to lock this concept in, a comparison. With capacitors, the voltage across the capacitor increases and the current decreases until, at 5 time constants:

V = supply voltage
dV/dT = 0
I = 0

With inductors, the current increases and the voltage across the inductor decreases until, at 5 time constants:

I = final value
dI/dT = 0
V = 0

Remember, we're fudging a little: 5L/R has been chosen as an arbitrary point where the voltages and current are “pretty much” stabilized.

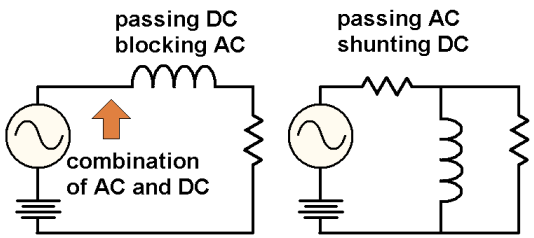
INDUCTORS AND ALTERNATING CURRENT

The higher the frequency, the closer the inductor comes to being an open circuit. Remember that an inductor has inertia, like a flywheel. Trying to quickly change the current through an inductor is like fanning yourself with a brick: the faster you try to move it back and forth, the smaller the movement. If this movement represents an AC signal, the higher the frequency, the less will get through; as frequency rises, an inductor increasingly resembles an open circuit. On the other hand, at lower frequencies an inductor increasingly resembles a piece of wire.

Thus the inductor is the opposite of a capacitor: it's a device that will pass DC but impede AC, increasingly blocking it as frequency rises. One common use is to remove stray RF from a power supply wire at the point where it enters a piece of equipment, so it won't get into the circuitry.

Another use you're familiar with: that cylindrical lump on a data cable is an inductor. It's there to attenuate the highest frequencies so the cable doesn't radiate excessive interference.

An inductor can also be used the other way, to short two points together at DC but allow AC to pass.

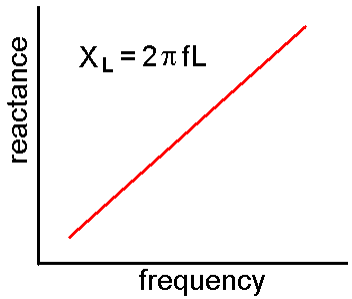


INDUCTIVE REACTANCE

Like a capacitor, an inductor ideally does not dissipate energy but stores and releases it; thus an inductor has reactance, specifically **inductive reactance**, $\mathbf{X_L}$, measured in Ohms. The formula is:

$$X_L = 2\pi fL$$

where X_L is inductive reactance in Ohms, f is frequency in Hz, and L is inductance in Henrys. Again, the 2π factor converts from angular velocity to frequency. Reactance is the electrical behavior caused by inductance, the actual quantity we deal with when calculating the amount of opposition to current flow.



Observe how similar the formula is to the reactance formula for capacitance. However, this time the "good stuff" is not in the denominator of a fraction. That's because inductive reactance has a *direct* relationship to frequency and inductance (reactance rises with inductance and frequency) as opposed to capacitors which have an *inverse* relationship.

As with capacitors, the voltage and current are 90 degrees out of phase, but with inductors the current lags the voltage by 90 degrees. This will be easy to remember if you keep in mind that an inductor has inertia, so current will always lag behind.

Real Henry-sized inductors are rare. Actual inductors tend to be sized in the millihenrys, microhenrys and nanohenrys. Related equations:

$X_L = 2x10^{-3}\pi fL$, where L is in millihenrys,
$X_L = 2x10^{-6}\pi fL$, where L is in microhenrys, and
$X_L = 2x10^{-9}\pi fL$, where L is in nanohenrys.

Let's work an example. Suppose we want to find X_L of a 1mH inductor at 10kHz. We start with the basic equation:

$$X_L = 2x10^{-3}\pi f$$

Note that I rearranged the terms, just cuz I felt like it. You remember the math principle that, in a string of multiplications, order doesn't matter.

Next we plug in the values we know, then crunch the numbers.

$$X_L = 2x10^{-3} \times 3.14 \times 10^4 \times 1$$
$$X_L = 2x10^1 \times 3.14$$
$$X_L = 6.3 \times 10^1$$
$X_L = 63$ (pretty close)

So, at 10kHz a 1mH inductor has a 63Ω of inductive reactance. Let's scale these quantities up and down.

The reactance of an inductor is directly proportional to its size; at 10kHz, a 2mH inductor has 126Ω of reactance, while a ½mH inductor has only 31.5Ω.

The reactance of an inductor is directly proportional to frequency; at 20kHz, a 1mH inductor has 126Ω of reactance, while at 5kHz it has 31.5Ω.

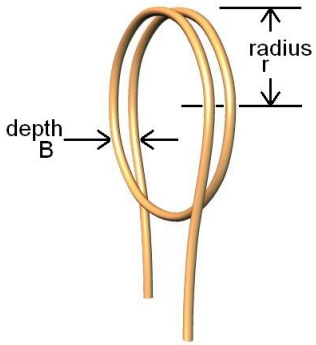
INDUCTORS IN SERIES AND PARALLEL

Inductive reactances in series and parallel work like resistors. With series inductors, total inductance is the sum of individual inductances. With parallel inductors, total inductance is found by the product over sum or reciprocal formulas, and is always lower than any of the individual inductors.

It might be surprising that X_L doesn't work the opposite way from X_C, but it doesn't; X_L, X_C and R all impede current and all obey the same series and parallel formulas. The difference is, the relationship between L and X_L is direct, while the relationship between C and X_C is inverse. That's where the opposites business comes in.

FACTORS THAT GOVERN INDUCTANCE IN AN AIR WOUND INDUCTOR

The diameter, length, number of turns and number of layers in an air-core coil determine the inductance. For a single-layer inductor,

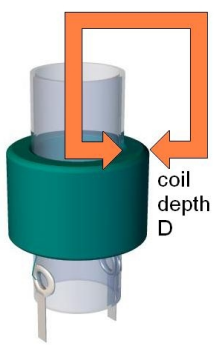


$$L = \frac{r^2N^2}{9r + 10B}$$

where r = radius of the coil in inches, N = number of turns, and B = length or depth of the coil in inches.

I'm being a little arbitrary; length is usually indicated not by B but by a lower-case L, but that can be confused with both I and 1. So I'm pulling rank.

The formula is most accurate for long, thin coils and less so for short, squat ones. For multilayer coils, where a winding is doubled up on itself may times like a spool of thread, the formula changes a bit:



$$L = \frac{0.8r^2N^2}{6r + 9B + 10d}$$

where d=depth or thickness of coil.

Inductance is very sensitive to the radius of the coil and the number of turns, because they're both square functions. Increasing either a little increases inductance a lot. It's also a bit sensitive to the length of the coil, which is affected by the wire gauge. Increasing the length decreases inductance.

FERROMAGNETIC CORE INDUCTORS

When an inductor is wound on a ferromagnetic core, and especially if the core forms a complete magnetic circuit, coil radius and the like are no longer the controlling factors. Inductance then is mostly governed by number of turns and by the size, shape and permeability of the core.

A core introduces so many extra terms into the equation that I'm not even going to try to describe how to calculate inductance. Partly because I couldn't do it myself. This is a challenge even for engineers, and not something a tech will be required to do.

However, there is one aspect of ferromagnetic inductors we have to discuss, and that is that there is a limit to the magnetization we can achieve. In chapter 5, I hinted at it when I mentioned saturable reactors.

SATURATION OF INDUCTORS

In EM1, we saw that there is a limit to the magnetism a ferromagnetic material can hold. The condition of full magnetization is called **saturation**. This is of great importance in inductors.

An air-core inductor cannot be saturated. But with ferromagnetic cores, it is quite possible for the core to become saturated at some point during an AC cycle – actually two points, the positive-going and negative-going voltage transitions.

Since current lags voltage by 90 degrees, the voltage zero-crossings are the points where current is highest, thus magnetization is highest and saturation most likely.

When the core saturates, the changing field stops changing, back voltage is no longer created, and reactance drops to near zero, leaving only the low DC resistance of the wire in the coil. In other words, the inductor becomes a near-short circuit twice each cycle. So we have to be sure we stay out of saturation. This is an engineer's problem, not yours, but you should understand why it occurs.

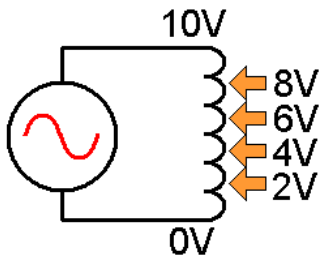
Saturation behavior depends primarily on three things: the AC voltage, the AC frequency, and how much iron there is to magnetize.

Saturation and Voltage

Regarding voltage, recall that it is voltage across the inductor that causes the magnetization to change. The higher the voltage, the faster the core magnetizes, and the more likely it is to hit saturation.

This doesn't mean high voltage inductors have to be big. Suppose we start with a certain voltage and a certain number of turns.

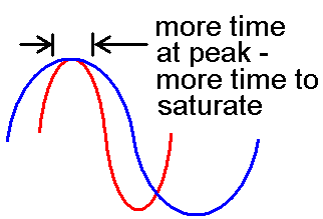
Now let's double the number of turns and also double the voltage. If you were to measure the voltage at many points along the coil, you would find that the voltage at each point is proportional to the fraction of the total number of turns at that point. At the midpoint of the coil, corresponding to the original number of turns, we'd find half the voltage, corresponding to the original voltage. In other words, from a saturation standpoint, we're back to where we started.



If putting more turns on an inductor lets it handle a higher voltage, couldn't we just pile on a gazillion turns and never have a problem? Well, when we were looking at air-core inductors in chapter 5, we already discussed why that won't work: lots of small-gauge wire means a resistance so high that it swamps the inductance. There are very few times in electronics where we get something for nothing; instead, we usually balance tradeoffs to obtain the best compromise.

Saturation and Frequency

The maximum voltage occurs at the peaks and valleys of a sine wave. The lower the frequency, the wider those peaks and valleys are, and the longer the period during each half cycle that the current is rising. That means, for a given applied AC voltage, peak current will be higher at lower frequencies. When current is higher, magnetization is higher, and the inductor is more likely to hit saturation on peaks.



Thus saturation is a low-frequency problem. What “low frequency” means in a given case depends primarily on how much ferromagnetic material is in the core and how many turns there are in the windings. Some inductors begin to saturate when the frequency gets below 100kHz; others are good down to 10Hz.

Saturation and Size

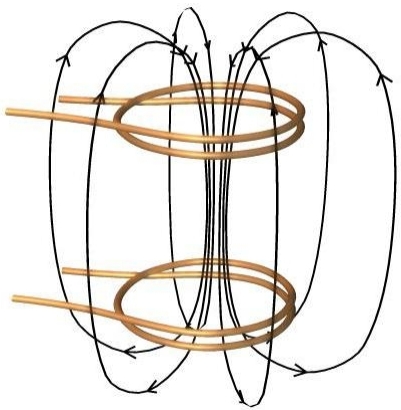
The only real way to solve a saturation problem is to put more iron in the core. That's why transformers that handle lots of power, like the ones that supply entire neighborhoods, are the size of cars. Size matters.

That's also why electric and electronic equipment in airplanes, where weight is important, have traditionally operated on 400Hz power, not 60Hz. Transformers and motors need a lot less iron when power is at a higher AC frequency. And it's why switching power supplies exist, as we will see in EM3.

MUTUAL INDUCTANCE

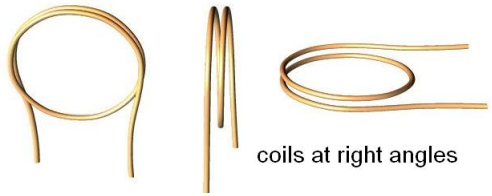
The general principles of inductance only hold true as long as no inductor invades any other inductor's "magnetic space." If inductors are close enough together that their fields are significantly shared, we get into something called **mutual inductance** and all bets are off.

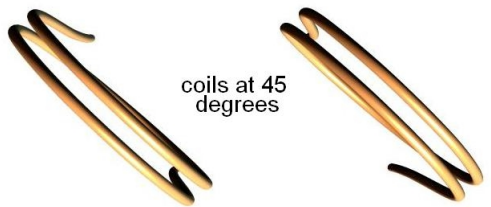
As we saw in chapter 5, **mutual inductance** occurs when two inductors are close enough to share a bit of mutual magnetic field, so some of the energy in one inductor is magnetically transferred to the other, transformer-style. This energy must be taken into account, and that brings whatever the second inductor is connected to into the equation.



This can get extremely complex, but it's enough for the tech to know it occurs and may have been intended to occur by the designer. Where mutual inductance is involved, the part of total inductance that is confined to one inductor is called **self inductance** to distinguish it from the part that is shared.

Mutual inductance can be a problem when it couples a signal from a place in the circuit where it's supposed to be to a place where it isn't. This can be avoided by turning inductors at right angles to each other or 45 degrees in the same direction. This creates a **null**, where part of the field cancels the other part.





Mutual inductance is, of course, the operating principle that transformers are based on. Also, a controlled amount of mutual inductance is a basic ingredient in the frequency-selective circuitry of radios and other wireless receiving devices. We'll be spending a fair amount of time on that in EM3 and EM6.

FARADAY'S LAW OF INDUCTANCE

You know the principle of Faraday's Law already, but you might as well be able to attach a name to it. In the form most useful to us, Faraday's law states:

"Induced voltage in a conductor is proportional to the *rate of change* of magnetic flux."

Flux, you will remember from EM1, is the strength of a magnetic field, described as the number of imaginary lines of force in a particular cross-sectional area. And flux is proportional to the current flowing in a coil. This gives us a satisfactory theoretical explanation for back EMF, which is the reason current and voltage act the way they do in an inductor, and for the fact that the output AC voltage of a transformer is a mirror image of the input voltage.

End of Chapter 7

EM2 Chapter 8
TRANSFORMER THEORY

Objectives of this chapter:

- *explain how a transformer operates*
- *define: winding, primary, secondary, core, turn, turns ratio*
- *define step-up and step-down and described how they are obtained.*
- *describe how loading of the secondary is reflected back through the transformer to become a load on the source*
- *describe taps, center taps, split winding*
- *describe transformer types, uses, construction*
- *define "stray impedance" and describe its effect on transformer operation*
- *discuss the matched-impedance environment in which transformer specifications are measured.*
- *describe the impedance ratios of step-up and step-down transformers*
- *define and discuss "sag" and how it relates to maximum loading*
- *discus current transformers and saturable reactors.*

Let's start with a recap of the information from chapter 5. A transformer:

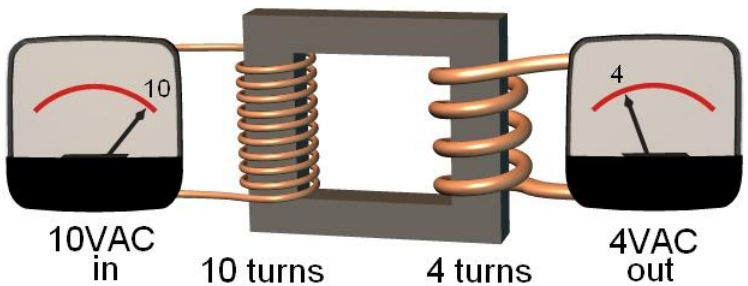
- works only with AC
- performs a double energy transformation: electricity to magnetism to electricity
- can transfer electrical energy from one place to another without an actual electrical connection
- can change the ratio of voltage to current

A transformer consists, at a minimum, of two coils of wire, magnetically coupled as closely as possible so as to have maximum **mutual inductance** – ideally, to share one magnetic field to such a degree that each individual coil has no field of its own. The transformer will also have a ferromagnetic core except at VHF and higher frequencies. The coils are called **windings**. The input winding is called the **primary** and the output winding is called the **secondary**. Each loop of wire in a winding is a **turn**.

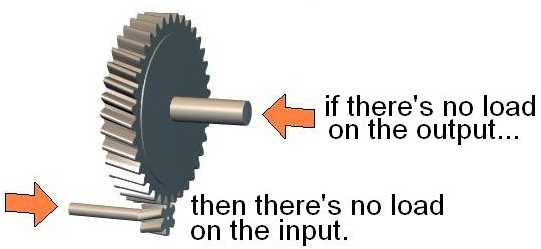
The output voltage can be different from the input voltage. If it is higher, the transformer is a **step-up** transformer; if lower, it's a **step-down** transformer.

What controls the voltage step up or down ratio is the ratio of the number of turns in the coils. This is called the **turns ratio** and, ignoring losses, the voltage ratio is the same as the turns ratio.

For example, in the illustration below, 10VAC into a 10-turn primary results in 4VAC out of a 4-turn secondary. This would be a high-power transformer operating at perhaps 200kHz and probably would have a ferrite core.
A 60Hz transformer would have the same turns ratio but hundreds of turns in each winding and a laminated iron core.

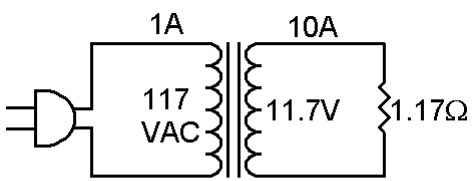


If the voltage is stepped up or down, what happens to current? In chapter 5, we saw that a transformer is like a gearbox. If the output shaft of the gearbox is not loaded, then there's no torque load on the motor connected to the input shaft (ignoring friction losses).



However, if you load the output, the gearbox reflects the load back to the motor, transformed by the gear ratio. If you have a 10:1 gear ratio and the motor runs at 10,000 RPM, the output will be 1,000 RPM. If you load that output at 10 foot-pounds of torque, the load back at the motor will be 1 foot-pound. Thus the ratio of speed to torque is changed, but rotating power, which is speed times torque, is neither increased nor reduced. Speed is stepped down, torque is stepped up.

The same thing happens with a transformer. An ideal transformer that's not connected to a load would not load the voltage source. For example, the voltage source might supply 117 volts to a 10:1 transformer, which would output 11.7V. If there is no load on the secondary, there will be no load on the source. But suppose we connect a 10-amp load to the output. (That would be 1.17 ohms, right? Very good.) Now the source has to supply 117 volts at 1 amp, which the transformer converts to 11.7 volts at 10 amps. In both cases the power level is 117 Watts.



10:1 voltage transformation
1:10 current transformation
power in = power out

I'm cheating a little. When there's no load on the secondary, the transformer is just an inductor. We know an inductor has reactance, and current will flow in the primary. But an inductor does not dissipate energy; it takes energy from the source during parts of the cycle and gives it back during other parts. You might compare this to a gearbox with a small flywheel. In any case, current flow through the reactance is insignificant compared to current when the transformer is loaded. That's a basic feature of a properly designed transformer: the load, not the transformer itself, is where the power goes. So we're justified in regarding the transformer as a device which passes power through but does not draw power itself.

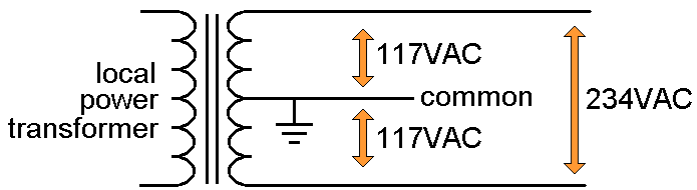
A reminder: transformers only work with AC. The output voltage is induced in the secondary by the *changing* magnetic field in the core. If you feed DC into a transformer, after a transient at the output you'll get nothing, except maybe smoke.

One of the reasons most countries switched from a DC to an AC power system in the 1920s is that it permits the use of transformers, which lets us use much higher generation and transmission voltages – thus lower current per watt – thus less power loss in the transmission lines. The 117VAC at your wall outlet comes into town at around 400kV.

Taps

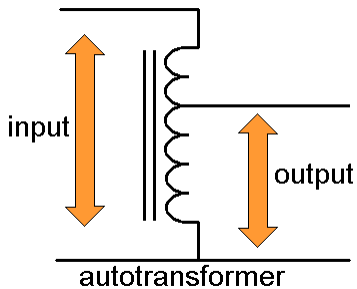
If you scrape a bit of insulation off an inductor winding and run a voltmeter up and down it, you will discover that the voltage at any point is proportional to the position in the winding. This is also true of transformer windings. It is often useful for a transformer to be able to offer a choice of input or output voltages. This is done by attaching **taps** at various places along the winding.

One of the most useful taps is a center tap. If we take this tap as the reference point, the ends of the winding provide two sources with the same voltage but opposite polarity.

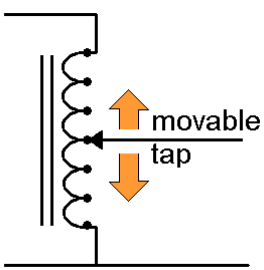


This is how the wiring in your house can provide 117 volts for lights and outlets and 234 volts (usually called 220 out of habit) for stove, electric water heater and central air conditioning. We get the 220 volts by connecting across the ends of the winding. We get 117 volts by connecting across either end and the center tap.

Autotransformers



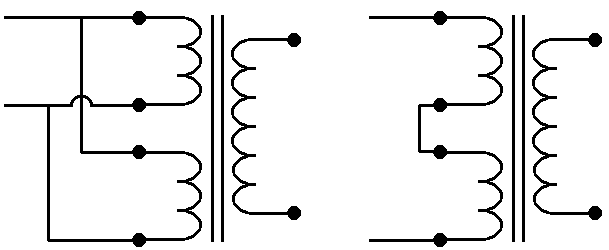
A tapped transformer where the input and output are on the same winding, and there is no second winding, is called an **autotransformer**. Its advantages are that more power can be handled for a given size, and it is simpler and less expensive to build. The disadvantage is that the isolation function is lost. Because there is no isolation, autotransformers cannot be used in most electronic equipment. However, they are extensively used in AC power distribution to change one voltage to another. Also, step-type speaker level controls are usually autotransformers with many taps.



Variable transformers, where the tap can be moved along the coil like the wiper of a potentiometer, were once widely used as lamp dimmers. In newer buildings, solid-state dimmers have long since taken over this function. However, variable transformers still find a place on technicians' workbenches, as it is often useful to be able to vary the AC voltage to equipment under test.

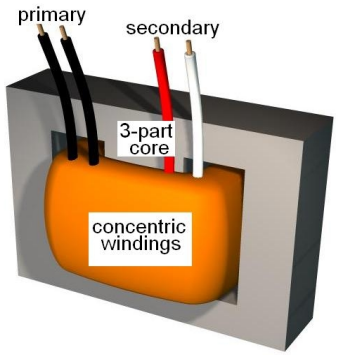
Multiple and Split Windings

There's no law that says a transformer can only have two windings. Complicated equipment often has transformers with several primary or secondary windings. For example, an old tube-type television will have a transformer with a 5V winding, a center-tapped 6.3V winding (identified by the nomenclature “6.3VCT”) and a center-tapped winding of hundreds of volts, and maybe one or two more. A CRT oscilloscope transformer requires a winding of hundreds if not thousands of volts, plus several other windings.



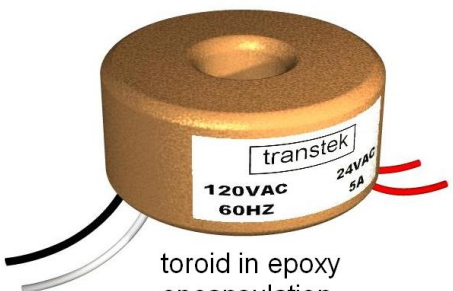
A single winding may also be split into two equal parts. The intention here is that they can either be wired in series, to connect to a specified voltage at a specified current, or in parallel, to connect to half the voltage at twice the current. An example would be a transformer made to connect to either 120V or 240V AC power.

Concentric Windings



Most transformers aren't actually built like the illustration on the first page. The core tends to have three “bars,” with primary and secondary windings stacked on top of each other on the center bar. The magnetic field travels out in both directions from the center bar through the side bars.

Toroid Transformers



toroid in epoxy encapsulation

If we take a very long strip of soft iron, coat it with plastic varnish, and wind it up, we'll have a laminated core in the shape of a donut. The proper name for that shape is a “torus” and transformers made this way are called **toroid** transformers. Their advantage is that they are more efficient magnetically, and their magnetic field is more completely contained inside the core; that is, they create less stray magnetism around them. A toroid transformer can also be smaller than its rectangular equivalent for a given power.

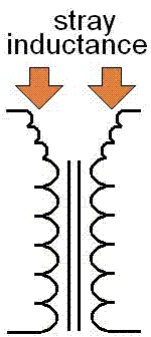
The disadvantage is that it's much more complicated to wind the coils on a toroid core. For this reason, most transformers are not toroids despite their advantages.

TRANSFORMER FREQUENCY RANGE

In theory, a transformer ought to behave the same at all AC frequencies. In practice, there is an upper and lower limit to its frequency range.

The low-frequency limit is governed by core saturation, just as with any iron-core inductive device. For a given power level, the lower in frequency you want a transformer to work, the more iron it has to have in the core.

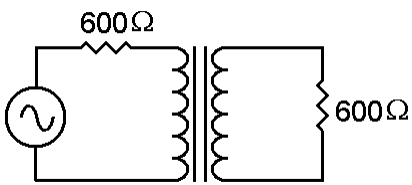
The high-frequency limit is constrained by the fact that not all the magnetic field created by the primary winding couples efficiently to the secondary – not all of it contributes to mutual inductance. Some winds up as a stray field, or self inductance, and this results in reactance that rises with frequency, interfering with transformer operation at high frequencies. Also, even with laminations, eddy currents cause core losses at high frequencies.



The bigger the core, the larger each turn of wire, and the more self-inductance it has. Thus good low-frequency operation works against good high-frequency operation and vice versa. For this reason, it's difficult to build transformers that work over a very wide frequency range, and transformers are usually built to handle only the frequency range they need to handle for the application.

SOURCE AND LOAD IMPEDANCES

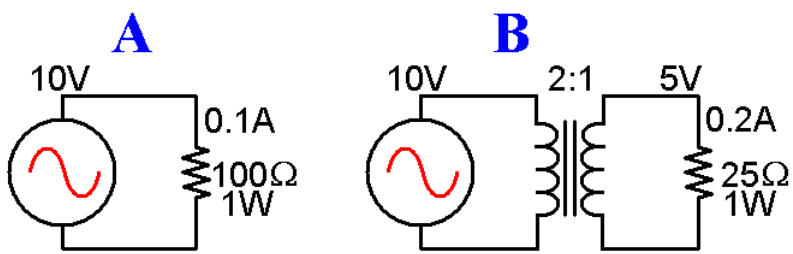
A signal transformer is designed to work in a **matched-impedance** environment. For a 1:1 transformer, it is designed to be driven by a source of a particular impedance, and loaded with the same impedance.



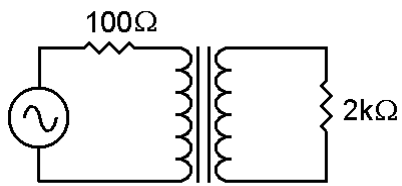
For a step-up or step-down transformer, *the matching impedances change with the **square** of the voltage ratio*. For example, if the transformer in the illustration above were a 2:1 stepdown unit, it would expect to see a 600Ω source and a 150Ω load. That is, the impedance ratio would be 4:1, or the voltage ratio squared.

This is best explained by example. Suppose we have a 10V source and a 100Ω load as in "A" above right.
10V / 100Ω = 0.1A of current. 10V x 0.1A = 1 watt of power.

Now suppose we insert a 2:1 transformer as at "B." A transformer does not make or, ideally, lose power; the output should still be 1 watt. A 2:1 stepdown means the output voltage will be cut in half, to 5V. At the same time, current will be doubled, to 0.2A. What load resistance does this call for? 5V / 0.2A = 25Ω. So a 2:1 voltage transformation results in a 4:1 impedance transformation.



There's no law that says a transformer must actually be used in a matched environment. In audio work especially, a 600Ω transformer is likely to be driven with a lower-impedance source and loaded with a higher-impedance load. This improves both low and high frequency response, and lowers distortion.

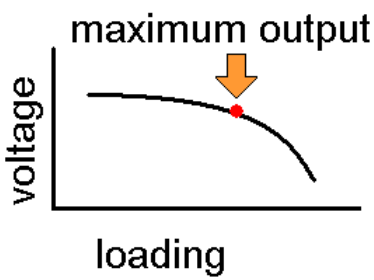


It's not a good idea to run a signal transformer with no load at all, because the inductance of the secondary and the capacitance of cables etc can cause a peak at the high end of the transformer's frequency range. This is called **resonant** behavior and we're going to study it in chapter 8.

With RF transformers, you don't have the option of fiddling with source and load impedances. RF distribution systems must almost always be a single impedance throughout, for reasons we'll discuss in chapter 10.

MAXIMUM POWER, VOLTAGE SAG AND HEAT

This loading business brings up something you could probably have guessed: there's a limit to the power you can get out of a transformer. If you try to get too much – that is, if you hang too low an impedance across the output – you'll get excessive voltage drop within the transformer itself, due to resistive and inductive losses. The secondary's DC resistance forms a voltage divider with the load. The primary resistance counts too; after all, it's in the circuit between the source and the load.



So, the more heavily you load the output, the more power you get, up to a point, but beyond that point, the *less* power you get as the voltage is pulled down. This is called **sag**. Where does that lost power go? Heat, right. A power transformer running at its specified maximum output gets quite warm; don't try to run a transformer beyond its limits.

With a signal transformer, losses within the transformer are called **insertion loss**. This is usually specified in decibels, e.g. a 1dB insertion loss means the signal will come out 1dB weaker than it went in.

RADIATION AND PICKUP OF STRAY FIELDS

Like any other inductive device, a transformer can radiate magnetism. Its field is never perfectly confined to its core, though as mentioned above, if its core is a toroid, radiation will be reduced.

One way to keep radiation from coupling into other circuitry is to pay attention to orientation. With a rectangular core, the stray field is strongest around the edges of the transformer. This should be placed at right angles to the most sensitive circuitry.

A second way is to maximize distance between the transformer and whatever it might interfere with. In EM1 we saw that a magnetic field's strength is inversely proportional to the square of distance, which makes distance a powerful interference-reducing tool.

The third way is to encase the transformer in a **magnetic shield** consisting of multiple layers of a special material called **mumetal** which is a very soft iron. This shield forms a magnetic low-reluctance path that shunts stray fields around the transformer. Shielding requirements are most stringent for low-frequency transformers; at RF frequencies a light-gauge mumetal enclosure will do.

Often, it's not worth bothering. An all-digital circuit isn't going to be affected by a little 60Hz hum, nor is a radio circuit operating at 50MHz. Whether shielding is necessary depends on the application.

Transformers can also pick up radiation. The same remedies work with incoming radiation as outgoing.

Some transformers also have an aluminum or copper electrostatic shield between primary and secondary windings to prevent electrostatic noise from being transferred from primary to secondary. This shield must be grounded to the unit's chassis to be effective.

TRANSFORMER CLASSIFICATIONS AND USES

A **power transformer** handles 60Hz AC power. Its purpose is to change the voltage from an AC power line to a different voltage needed to power equipment, to a branch lighting or outlet circuit in a building, or to other transformers downstream, as in a high-voltage utility substation. An individual piece of electronic equipment usually has a power transformer in it, to change 117V to whatever the internal circuitry requires. The electrical isolation between the primary and secondary must be very dependable, as this is a life safety issue. A power transformer may include a thermal fuse which opens if the transformer gets too hot. This is buried deep within the transformer and is not replaceable; if a transformer ever gets hot enough to melt this fuse, it isn't safe to use any more.

An **audio transformer** handles audio frequencies. The range of human hearing is about 16Hz to 20kHz, but few audio applications require this entire band. A high-quality audio transformer will operate well from 30Hz to 30kHz. It is not easy to build a transformer to work well over this 1,000:1 range. In addition, protection against stray magnetic fields requires that the transformer be encased in a mumetal shield, perhaps several layers of it. Due to these factors, a top-quality audio transformer costs upwards of $100.

Voice communication only requires a range of 300Hz to 3kHz, though this will yield a nasal, telephone-like sound.

A **modem transformer** is an audio transformer with poor low frequency response. There are no frequencies below 200Hz in modem data, so the transformer doesn't need a lot of iron. Magnetic shielding is not required, but many modem transformers have an electrostatic shield between primary and secondary.

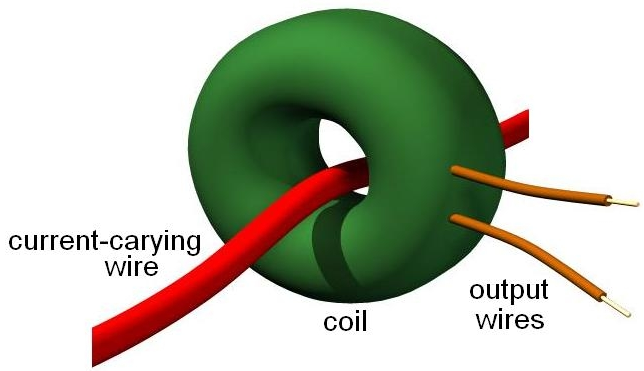
Switching power transformers are high-power transformers built to operate at 50kHz or higher. We will be dealing with switching power supplies in EM3.

RF transformers are made to cover many ranges, from a few hundred kHz to hundreds of MHz. At higher frequencies, they may be air-core units. **Antenna** and **cable transformers** are RF transformers.

Current Transformers

A current transformer is not intended to transfer significant power from primary to secondary. It is intended to siphon off just enough power – we're talking microwatts – to allow sensing circuitry to measure current without having to break the circuit and insert an ammeter.

In a current transformer, the primary usually consists of a single heavy wire, passing through a coil which forms the secondary. Unlike a regular transformer, in a current transformer, the voltage drop across the primary is very small. The coil, with many turns, steps the small induced voltage up to the point where it can easily be measured. This does away with shunts, power dissipation and possible bad connections in a high-current circuit.

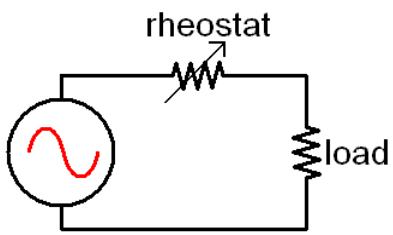


Current transformers are used in **ground-fault interrupt** or **GFI** outlets. In this case, the primary consists of two wires – the hot and common of an AC power circuit. Current going to the outlet on one wire ought to be precisely equal to current coming back on the other; thus the two magnetic fields should cancel and the output of the transformer should be zero. If it isn't, that means a fault has occurred and some current is returning to ground via another path, e.g. a person's body. The imbalance causes the current transformer to output a voltage, which trips the breaker.

Saturable Reactors

A saturable reactor has two windings, but strictly speaking, it isn't a transformer – it is not intended to transfer power from primary to secondary. Its purpose is to substitute for a big resistor.

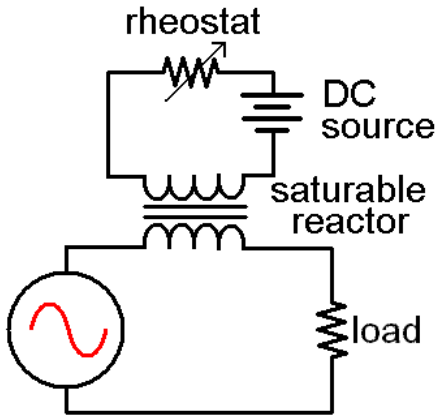
It is possible to control the voltage to a large load with a large series rheostat. At one time, this was a common occurrence. Unfortunately, it isn't very efficient; the rheostat burns up a lot of power. A dimmer for a church sanctuary, for example, was typically two feet in diameter, and you didn't dare install one in a closet for fear of starting a fire. In England, resistive dimmers consisting of earthenware jugs of salt water were used to control stage lighting. The water would get hot and evaporate, and have to be topped off frequently.



A series reactor, on the other hand, dissipates comparatively little power. Reactance is controlled by feeding a DC current through a second winding, deliberately pushing the core into saturation. This dramatically lowers X_L.

The main difference from a transformer is that it must be constructed with dual, cross-connected AC windings or other techniques to avoid coupling a lot of AC power into the DC circuit.

Saturable reactors have largely been replaced by solid-state circuitry, but hey are still used in some switching power supplies and arc welders.



End of Chapter 8

EM2 Chapter 9
RC, RL, RCL PRINCIPLES

Objectives of this chapter

- *evaluate combinations of resistance, capacitance and inductance in terms of impedance, voltage, and phase*
- *describe how various RC, RL and RLC combinations respond to different frequencies*
- *define and discuss resonance, Q and damping*
- *define frequency response*
- *discuss in general terms the circuits and frequency responses of the most common types of filters*

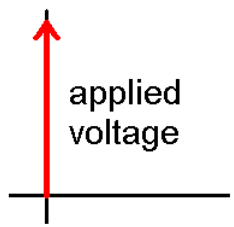
Now that we've looked at capacitors and inductors, it's time to look at combinations of these components:

- **RC** circuits, combining **resistors and capacitors**
- **RL** circuits, combining **resistors and inductors**
- **LC** circuits, combining **inductors and capacitors**
- **RLC** circuits, which combine all three.

Combining these components means we have more work to do, but also that we *can* do more. Combinations of resistors, capacitors and inductors give us tools to do things we couldn't begin to do before.

For the first part of our discussion,we'll assume we're dealing with sine waves, one frequency at a time. We'll also assume that the AC voltage source supplying these sine waves has essentially zero impedance and any load that may be connected has essentially infinite impedance, so we can ignore them in our calculations. This is often the actual case in the real world.

PHASOR REPRESENTATION OF VOLTAGE DROPS



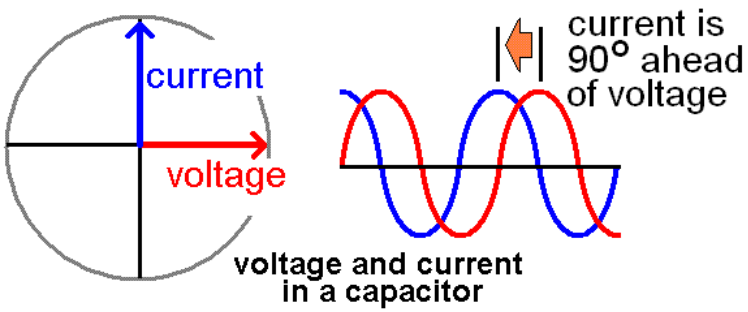
It is convenient to represent the AC voltage applied to a circuit as a phasor pointing straight up, corresponding to a sine wave at its peak. A voltage divider can then be represented as a series of phasors pointing the other way, dropping the voltage back to zero. (Remember Kirchoff – the voltages around a loop add up to zero.)

SERIES RC COMBINATIONS

When we connect an AC voltage across a series RC string, the RC string divides the voltage just like a resistive divider – but the voltages across the resistor and capacitor add up to more than the applied voltage!

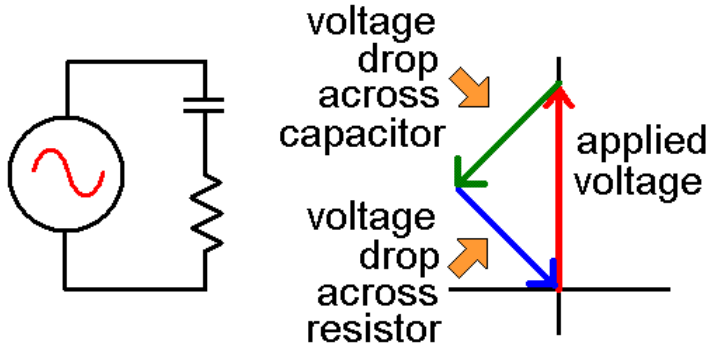
That would seem to be impossible, but it's true. The reason can be found in two facts:

- Voltage and current in a resistor are **in phase**, meaning that current is precisely in step with voltage.
- Voltage and current in a capacitor are **90 degrees out of phase**, with current leading voltage.

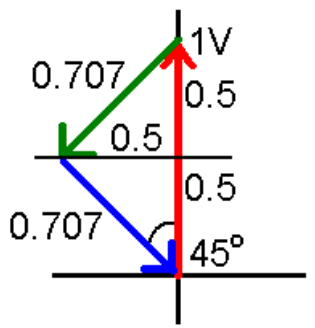


voltage and current in a capacitor

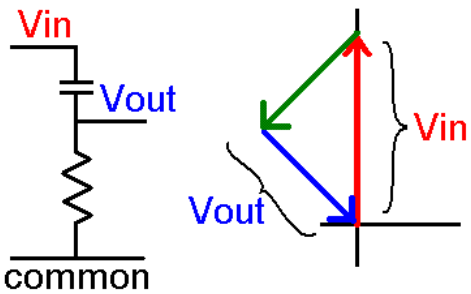
What this means for a series combination of resistor and capacitor is that there is a 90-degree phase difference between the voltages across them. Let's look at a frequency where R and X_C are equal, which means as you'd expect that the voltage drops across R and C will be equal. We'll draw the applied AC as a red vertical phasor, showing the voltage at its peak. If we return that voltage to ground through two equal voltage drops, with a 90-degree angle between them, we'll get the red-green-blue right triangle shown above right.



Time to get out the calculator. We want to find the length of the green and blue lines. Working from the angle at the bottom, the voltage drops across the capacitor and resistor are the opposite and adjacent. Since they are equal, the opposite divided by the adjacent – the tangent – is 1. Punch it in and the calculator will tell you this is a 45-degree angle. The sine and cosine of a 45-degree angle are both 0.707, meaning that if the hypotenuse, the red phasor, represents a peak of 1V, then the other sides of the triangle, the green and blue phasors, are 0.707V and 0.707V, which sum to 1.414V. You could say the voltage takes a detour on its way to ground, thus the sum of the voltage drops is larger than the original voltage.

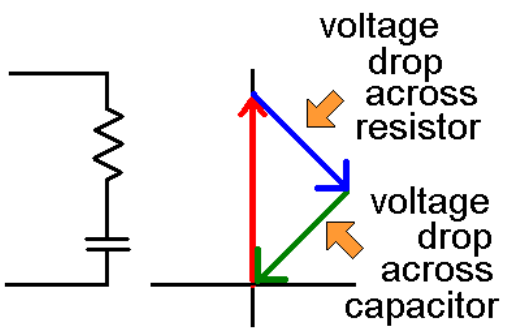


Let's continue the voltage divider analysis. In a voltage divider, the input is applied across the series string and the output is taken across the lower element, the shunt, which in this case is the resistor.

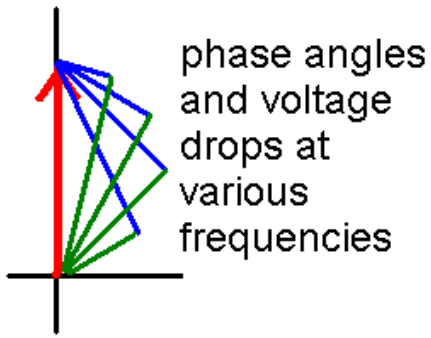


We can see that the triangle point representing the junction between capacitor and resistor is to the left of the input phasor. We already know that the angle at the bottom, next to the origin, is 45 degrees, and phasor rotation is counter-clockwise, so the output leads the input by 45 degrees, and its amplitude is 0.707 times the input voltage. In effect, the stored charge in the capacitor pushes the output voltage out in front of the input voltage, like a person pushing a lawnmower around a circle.

What if we reverse the components? Now we have a triangle that extends to the right; the output voltage lags the input voltage. The phase relationship in the capacitor hasn't changed, but now that it's shunting the output instead of in series with it, the time it takes to charge, discharge and reverse-charge the capacitor every cycle delays the phase. Instead of current leading voltage, we now have voltage lagging current, which is the same thing stated a different way. Think of a reluctant dog being pulled along by a leash.

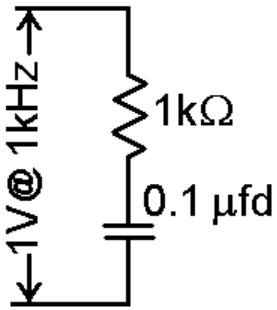


Bear in mind that this equal voltage division only occurs at one frequency. At any other frequency, X_C will be different, so the voltage ratios will work out differently. The one thing that stays the same is that the angle between the R side of the triangle and the X_C side will always be 90 degrees.



With that bit of information, and your calculator's ability to calculate sines, cosines, tangents and angles, you can figure out the voltage divider ratio and phase angle for any frequency. The ratio of X_C to R at a particular frequency amounts to the tangent of some angle, which gives you an entry point into the calculation.

Just for practice, let's take a frequency of 1kHz, a series R of 1kΩ and a shunt 0.1ufd capacitor and calculate the voltage drops and phase angles. The first thing is to find X_C of a 0.1ufd capacitor at 1kHz.



You remember the general formula:

$$X_C = 10^6 / 2\pi fC$$ (C is in microfarads)

Substituting the quantities we know,

$$X_C = 10^6 / 6.3 \times 10^3 \times 10^{-1}$$
(inserting "x" to avoid a confusion – do what ya gotta)

Then we cancel exponents:

$$X_C = 10^4 / 6.3$$

...and do the division:

$$X_C = 1.67k\Omega$$ (pretty close)

This gives us a divider with R=1kΩ and X_C = 1.67kΩ, meaning the voltage drops will have a ratio of 1:1.67, with the larger voltage drop across the capacitor. Those will not be the actual voltage drops, but the *ratio* of voltage drops.

The tangent of the angle next to the origin is 1 / 1.67, which is almost exactly 0.6. Hey, calculator – what angle corresponds to a tangent of 0.6? 31 degrees, thank you very much. And what is the sine of 31 degrees? About 0.52, thank you again.

The sine is the opposite divided by the hypotenuse. In this case, the hypotenuse is the input voltage, which is 1V. So we multiply the input voltage by the sine:

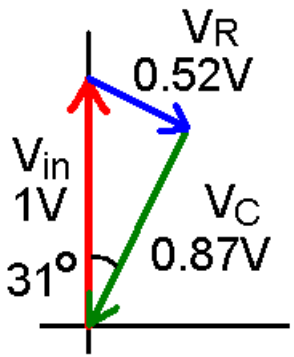
$$V_R = 1V \times 0.52 = 0.52V$$

... to get the voltage drop across the resistor, 0.52V.

We could look up the cosine to find the voltage drop across the capacitor, but we already know it's 1.67 x the voltage across the resistor:

$$V_C = 1.67V_R = 1.67 \times 0.52 = 0.87V$$
(pretty close)

So, the voltage drop across the resistor is 0.52V and the drop across the capacitor is 0.87V. Voltage at the junction of R and C lags the input by 31°.



There's another way of doing this that you may find simpler: just do the Pythagorean calculation, then scale.

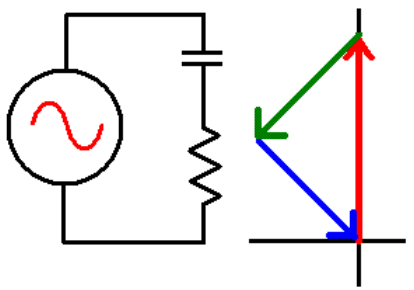
$1^2 = 1$ (opposite)
$1.67^2 = 2.79$ (adjacent)
$1 + 2.79 = 3.79$ (sum of sq.)
$3.79^{0.5} = 1.95$ (hypotenuse)

This gives us a triangle where the sides are in the correct ratios: 1, 1.67, 1.95. Next, we scale. We find the reciprocal of 1.95, which is just a matter of punching it into the calculator and hitting the "1/X" button: 1 / 1.95 = **0.514** That is, a multiplier of 0.514 is what it will take to shrink the hypotenuse, 1.95 back to 1, which is the input voltage.

Now we need to shrink the other sides. 1 x 0.514 is of course 0.514, the voltage drop across the resistor. 1.67 x 0.514 is 0.858, the voltage drop across the capacitor.

Note that I've been careless enough that the numbers don't match perfectly. This kind of fudging drives a lot of people up the wall, but please remember that real-world component values aren't precise either. If these are typical 5% parts, the actual numbers could easily be 0.55 and 0.79 or 0.46 and 0.93. There are cases where the ratios must be precise, in which case you'll be working with expensive special parts. Then you will be justified in doing the math more precisely than I have.

Hypothetical question: in the circuit at right, what is the phase relationship between I_C, current through the capacitor, and I_R, current through the resistor?

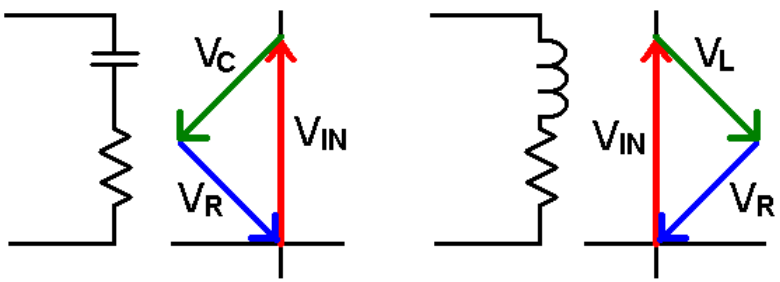


Nope! They're in phase! They have to be; it's a series circuit and current in a series circuit is the same at every point at every instant. The current through both components is the same, and it leads the input voltage by 45 degrees. Note that this agrees with what we know about resistors: in this circuit, V_R leads V_{IN} by 45 degrees, and current through a resistor is in phase with voltage across it. Another instance of what I've said before: finding an answer is largely a matter of finding the right way to look at the question.

SERIES RL COMBINATIONS

This is going to be short and sweet. Series RL circuits work exactly like series RC circuits except that X_C is inversely proportional to frequency while X_L is directly proportional to frequency, and the voltage/current phase relationships are reversed. In a capacitor, current leads voltage by 90 degrees; in an inductor, current lags by 90 degrees.

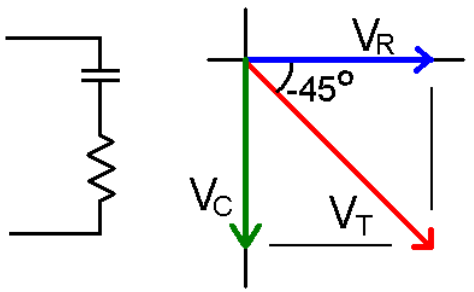
The math you need is the math you've already learned for RC circuits. For example, in an R_C or R_L circuit as shown below, with an input voltage of 1V and X_C or X_L equal to R, the voltage drops will be equal, but in the RC circuit the output voltage will lead the input voltage by 45 degrees, while in the RL circuit it will lag by 45 degrees.



A DIFFERENT WAY OF SHOWING VOLTAGES

We've shown the voltage drops of the resistor and reactor as fractions of the total voltage. It's often useful to turn things around and show the total voltage as the sum of resistor and reactor voltages. If we redraw the phasor diagram with V_R along the X-axis, horizontally, then V_C will be vertical (cuz it's 90 degrees from V_R, right?).

In the circuit at right, V_C lags V_R by 90 degrees. So, if we put V_R on the X-axis, the green phasor representing V_C goes downward on the Y-axis.



Next, we note the lengths of the X and Y phasors and draw the applied voltage phasor V_T having tip coordinates of X, Y.

We could show V_R extending to the right from the tip of V_C, which would give us a triangle as before, but this gives the same result and shows how two voltages sum. It's the standard method of showing a voltage that's the algebraic sum of two 90-degree voltages, V_{TOTAL} or V_T.

Mathematically, if we know V_R and V_C, we simply square them, add them and take the square root to get V_T. And V_R / V_C gives us the tangent of the angle, which in this case is -45^O.

Note the minus sign. Since phasors are thought of as rotating clockwise, and we've put V_R on the X-axis, which is zero degrees, that means V_T is 45 degrees behind or -45°.

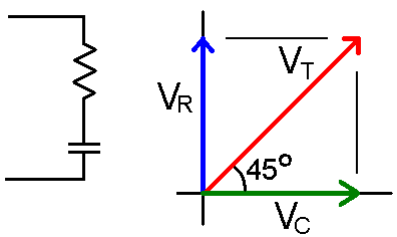
Let's say V_R and V_C are 1V. Doing the math, that makes V_T 1.4V. (Square root of 2. Yeah, you caught that.) So we could notate V_T as:

$$\mathbf{V_T = 1.4V, -45^o}$$

or, depending on what's more useful in a given situation:

$$\mathbf{V_T = 1V, -j1V}$$

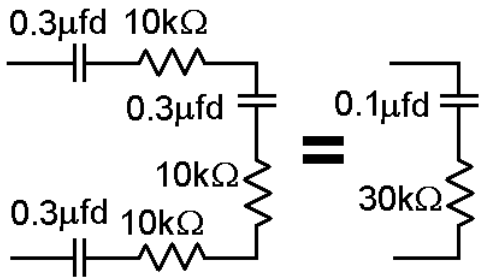
If we reverse R and C, the output is across C, V_T will lead the output by 45°. The same type of diagram can be used for R_L combinations, as long as you remember to switch the phasors so V_L leads V_R.



Phasor diagrams are naturally suited to showing peak voltages. Once the various ratios have been figured out, it's easy enough to convert to RMS: multiply any of the peak voltages by 0.707.

LUMPING VALUES

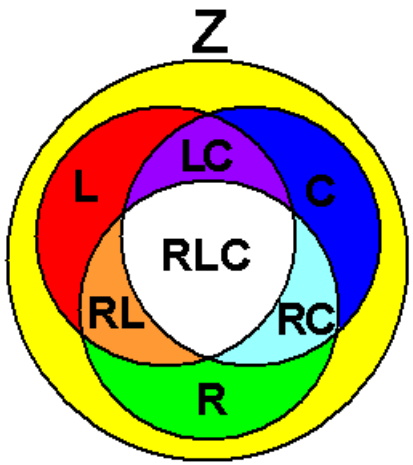
When a series circuit contains several resistors and capacitors, for purposes of finding voltage drops and phase angles, all the resistances can be lumped together into one resistance and all the capacitors can be lumped together into one capacitance. Remember that the total capacitance of capacitors in series is lower than any individual capacitance, and must be found by product-over-sum or reciprocal method.



IMPEDANCE

I've mentioned this before, but it's worth repeating. We know that when we're dealing with resistance, the current and voltage are in phase, and with reactance, they're 90 degrees out of phase. When resistors and capacitors are connected in series (or parallel for that matter), the phase can be anywhere from +90 to -90 degrees, and it will change with frequency.

In more complicated circuits, phase shift can even exceed 90 degrees. We need a word to cover every case. That word is **impedance**, abbreviated **Z**. Like resistance and reactance, it opposes the flow of current and is measured in Ohms.



An impedance may be pure resistance with no reactance component, just as a color monitor will display black-and-white text without color. An impedance may be all reactance. Most of the time, it's a combination of resistance and reactance.

When impedance is given as a simple number, there's an implied fudge factor. Anything that has capacitors and inductors in it, or capacitive and inductive effects, is likely to have an impedance that changes with frequency. Often, a stated impedance is just a target, a number that's approximately valid across a certain portion of the frequency spectrum.

For example, the 600Ω output of an audio mixing console, if it's really 600Ω (most aren't) is almost pure resistance. On the other hand, a 600Ω telephone line has all sorts of inductance and capacitance characteristics distributed along its length, and is only approximately a resistive 600Ω line over a limited range of frequencies.

So, impedance means “opposition to current flow” – including *all* combinations of resistance, capacitance and inductance.

In calculations, Z takes the place of R:

$$V = I \times Z$$
$$I = V / Z$$
$$Z = V / I$$

In diagrams, the separate R, X_L and X_C components of Z may be shown, or they may be represented by a resistor. If the resistor is labeled “Z,” assume that the resistor may represent a complex impedance with X_C and X_L.

IMPEDANCE OF A SERIES RC OR RL CIRCUIT

When two resistors are connected in series, their resistances add; the same thing happens with two capacitive reactances or two inductive reactances in series. Summing a resistance and a reactance is just a little more complicated.

The formula for a series resistance and reactance is

$$Z^2 = R^2 + X^2$$

Obviously, that's just the Pythagorean theorem for right triangles: the squares of the short sides, summed, equal the square of the hypotenuse. In this case, the short sides are R and X; the hypotenuse is the total impedance.

Since the squares of the two voltage drops add up to the square of the input voltage, naturally the squares of the impedances would add up to the square of the total impedance. To get the impedance, we square-root each side of the equation:

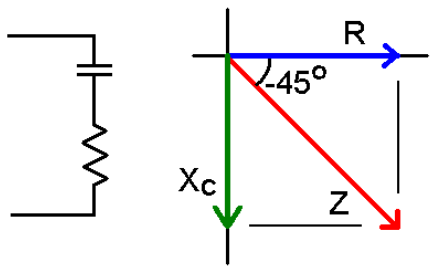
$$Z = \sqrt{R^2 + X^2}$$

Example: if you have a 10kohm resistor and 0.016ufd capacitor in series, and you have figured out that a 0.016uf cap has an X_C of 10kΩ at 1khz (there, I just saved you some work), then you can find the series impedance at 1kHz:

$$Z^2 = (10^4)^2 + (10^4)^2$$
$$Z^2 = 10^8 + 10^8$$
$$Z^2 = 2 \times 10^8$$
$$Z = 1.4 \times 10^4 = 14k\Omega$$

The exact same formula is used to find the impedance when the series combination is an inductor and resistor.

Remember how useful I said the alternate way of showing phasor addition could be? We can show R, X_C, Z and phase angle on the same kind of diagram. Note that Z is lower than it would have been if both components were resistors.



This has to do with the difference in phase angles, and the fact that the capacitor stores energy and gives it back, instead of dissipating it as a resistor would.

CURRENT IN A SERIES RC OR RL CIRCUIT

Now that we've found V_T (V_{TOTAL}, the input voltage) and Z, finding I will be simple. We use the modified Ohm's Law:

$$I = V_T / Z$$

In our example, suppose V_T is 1.4V and Z is 14kΩ or 1.4 x $10^4\Omega$.

$$I = 1.4 / (1.4 \times 10^4) = 10^{-4} = 100\mu A$$

Current leads voltage by 45°. If the reactance had been inductive, phase would have lagged. Note that we don't have to separately calculate I_R, I_C, I_T because, as noted earlier, this is a series circuit; there is only one current and it is the same everywhere.

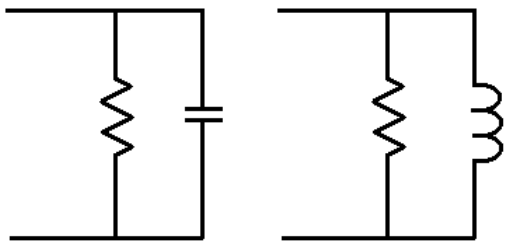
POWER IN A SERIES RL OR RC CIRCUIT

When we talk about power in a circuit with a reactance in it, we have to specify whether we mean **real power** or **apparent power**.

Apparent power is found by multiplying V_T by current, or dividing V_T^2 by Z. That gives us the power *in* the circuit, but not the power *dissipated* by the circuit. The portion of the power that's in the reactive element at any point in a cycle is not dissipated, but stored, and it's given back at other points in the cycle. So the only real power drawn by the circuit is the power in the resistor, which is found by multiplying the voltage drop across the resistor times current.

PARALLEL RC CIRCUITS

In a parallel circuit, voltage is the same across both components. V_R, V_C and V_T are all equal, so there is no need for a phasor voltage diagram. However, there will be separate I_R, I_C and I_T, as well as R, X_C and Z.

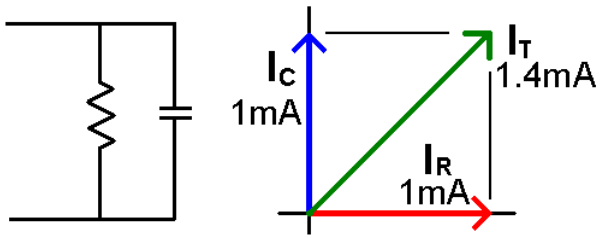


CURRENT IN A PARALLEL RC CIRCUIT

In a series circuit, you find Z and calculate I. In a parallel circuit, it's easier to find I first and calculate Z. Assuming you know R, and can calculate the value of X_C, the equation for parallel current follows the Pythagorean form:

$$I_T^2 = I_R^2 + I_C^2$$
$$I_T = (I_R^2 + I_C^2)^{0.5}$$

For example, if V is 1V, R is 1kΩ and X_C is also 1kΩ, then I_R and I_C will both be 1mA. I_R will be in phase with the applied voltage and I_C will lead by 90 degrees.



If you go through the calculations (as you know very dang well I want you to), you'll find that I_T is 1.4mA and leads voltage by 45°.

IMPEDANCE AND POWER IN PARALLEL RC CIRCUITS

Knowing the voltage, and having found the current, it's easy to find the parallel impedance:

$$\mathbf{Z = V / I}$$

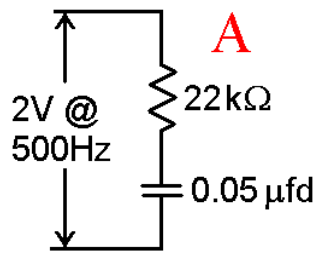
Apparent power is voltage times total current, or voltage squared over Z. Real power is voltage times I_R, or voltage squared over R. Total impedance will always be lower than the impedance of the individual components.

CURRENT, Z AND POWER IN PARALLEL RL CIRCUITS

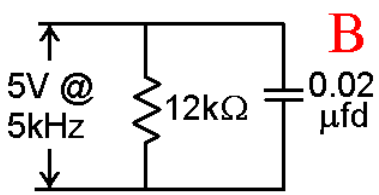
As you'd expect, these are the same as in RC circuits. The difference is, current lags voltage instead of leading it. Short, sweet, simple. Isn't it nice when that happens?

EXERCISES

Let's check and see if you've got the concepts to this point.



1. In figure A at right, find the impedance of the series combination.
2. What is the phase angle of V_C compared to V_T?
3. What are V_R and V_C?
4. In figure B at right, find the current in each component and the total current.
5. What is the total impedance of this circuit?
6. What is the real power dissipation of this circuit?



Answers:

1. First, find X_C (truncate to 3 places):

$$Xc = \frac{10^6}{2\pi fC}$$

$$Xc = \frac{10^6}{6.3 \text{ x } 5 \text{ x } 10^2 \text{ x } 5 \text{ x } 10^{-2}}$$

$$Xc = \frac{10^6}{6.3 \text{ x } 25}$$

$$Xc = \frac{10^6}{157.5}$$

$$Xc = 6{,}350$$

Then find total Z:

$$6.35^2 = 40.3$$
$$22^2 = 484$$
$$484 + 40.3 = 524$$
$$\sqrt{524} = 22.9$$

These numbers are in kilohms, so Z= 22.9kΩ. It may seem impossible that 22k and 6k in series could add up to only 22.9k, but if you draw the phasors you'll see that the angle is so great that X_C contributes very little to Z.

2. The tangent of the angle is R / X_C, 22 / 6.35 = 3.46. The angle is about 73.9 degrees.

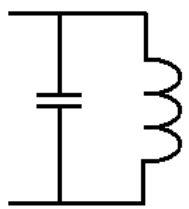
3. Find the hypotenuse: 22^2 = 484. 6.35^2 = 40.3. 484 + 40.3 = 524. (note truncation.) $524^{1/2}$ = 22.9. Scale: 22.9 / 2 = 11.45. V_R = 22 / 11.45 = 1.92V. V_C = 6.35 / 11.45 = 0.555V.

4. I_C = 3.14mA, I_R = 0.417mA. I_T = 3.56mA.

5. 1.4kΩ

6. 2.08mW

INDUCTANCE-CAPACITANCE COMBINATIONS



LC combinations are more complicated than RC or RL, but as in so many other cases, along with new complications come new tools that give us ways to do new things.

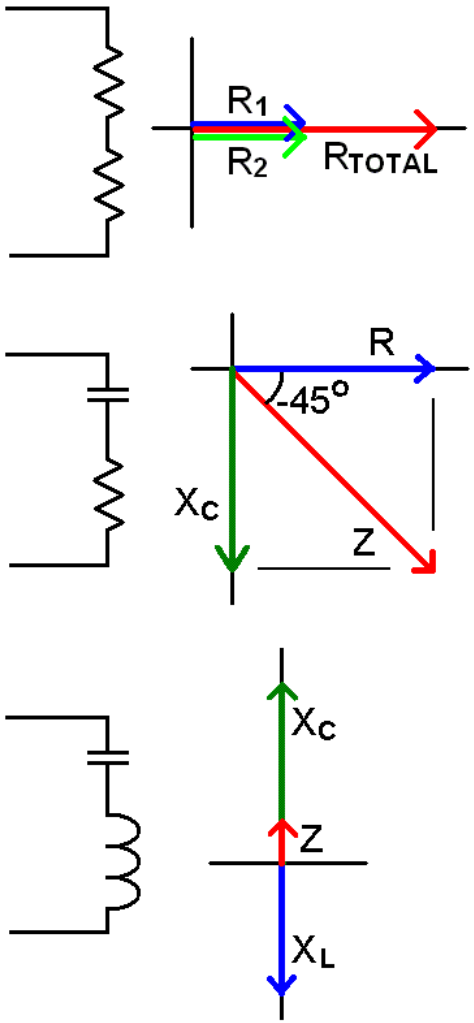
Strictly speaking, there is no such thing as an LC circuit, totally without resistance. As you know, there's resistance in everything, and you can't get rid of it, like death and taxes. However, it's convenient to think of ideal components when you're learning the principles involved. Moreover, in many real-world circuits, resistance amounts to 1% or less of total impedance, so theoretical circuit behavior is a pretty good model for real circuit behavior.

SERIES LC IMPEDANCE

Two series resistances add; series resistance and reactance partially add. But X_C and X_L, being opposite in phase, subtract. In a series LC circuit, the smaller quantity is subtracted from the larger quantity, and Z is whatever remains. In a series LC circuit, Z is always smaller than either X_C or X_L.

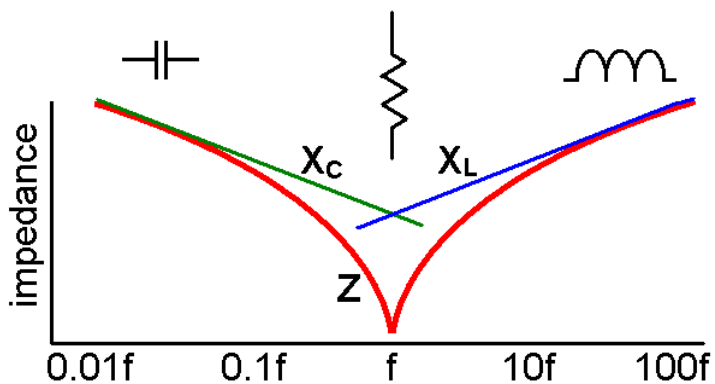
This would mean, at the frequency where $X_C = -X_L$, Z will be zero. If L and C were ideal components, that would be true.

This is often a tough thing for people to wrap their heads around. We're used to thinking of series-connected impedances as adding, but series-connected X_C and X_L subtract. This is because, as shown in the diagram at lower left, there are complimentary 90-degree phase shifts involved. In calculations, we put a negative sign in front of X_L to show that it subtracts from X_C.



I'm going to identify the frequency where $X_C = X_L$ by a lower-case **f**, for reasons that will soon become apparent. Let's look at how a series LC circuit's impedance behaves at f and above and below f.

At very low frequencies, X_L is so much lower than X_C that the inductor could just as well be a piece of wire. To the outside world, the series combination looks like the capacitor alone, and at low frequencies, it's nearly an open circuit. Z essentially = X_C, and it will be very high.

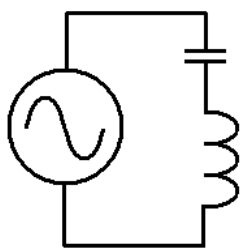


As frequency rises, X_C falls and X_L rises linearly, as you'd expect. At some point, X_L becomes large enough to matter. Since X_L subtracts from X_C, the impedance curve deviates downward. To the outside world, the circuit still looks like a capacitor, but a smaller one.

Let's go to the other end of the spectrum. At very high frequencies, X_C is so much lower than X_L that the capacitor might as well be a piece of wire. To the outside world, the series combination looks like the inductor alone, and at high frequencies, it's nearly an open circuit. Z essentially = X_L, and it will be very high.

If we reduce the frequency, X_L falls linearly; at some point X_C becomes large enough to matter and the curve deviates downward. To the outside world, the circuit still looks like an inductor, but a smaller one.

Near f, X_C and X_L approach equality, nearly canceling each other, and Z plunges. At f, Z = 0.



With real-world components, the capacitor and inductor have some resistance, called **parasitic** resistance. At f, this resistance constitutes the whole of Z. To the outside world, the series combination looks like a small resistor.

SERIES LC VOLTAGE AND CURRENT

At any frequency above or below f, V is the applied voltage and I is found by finding net reactance, $X_C - X_L$ or $X_L - X_C$, and applying Ohm's Law.

If we think of ideal components, the situation at f is impossible, a logical contradiction: a stiff voltage source that cannot be shorted out connected to a short circuit. This is one of those impossible cases I mentioned in the math chapter, where we encounter an equation where we're dividing by zero. In the real world, current will be determined by V and the total of all the small resistances in the system. This is a simple in-phase, V/I/R Ohm's Law problem.

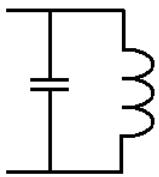
But the situation is much different with respect to the individual components. Imagine that we have a capacitor and inductor whose reactance at f is 100Ω, and there is 1Ω of resistance in the system. This is a typical real-world case with high-quality parts.

Current will be 1V / 1Ω = 1A. Since this is a series circuit, I_C and I_L will also be 1A. If the reactances are each 100Ω, then the voltages across each of them will be

$$1A \times 100\Omega = 100V$$

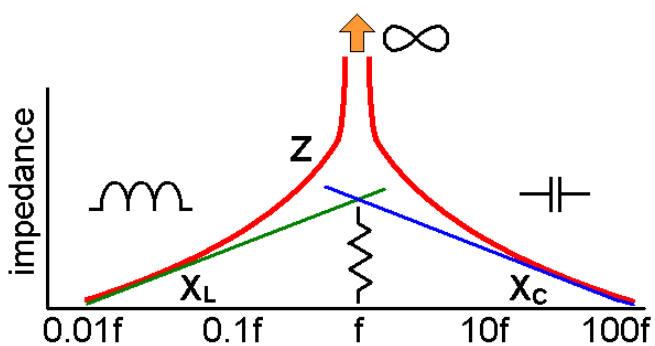
This is a startling result but it's quite real. It's a characteristic of series-connected LC circuits: the voltages across L and C, and the voltage with respect to ground at the junction between them, may be many times higher than the input voltage that created them.

What if we don't want the voltages to be that high? We reduce the applied voltage or add more resistance. We'll return to this and study it in more detail.



PARALLEL LC IMPEDANCE

When an inductor and capacitor are connected in parallel, the resulting circuit is often called a **tank**, probably because it can temporarily store energy. In a parallel LC circuit, the situation is opposite from the series LC circuit: Z is always higher than X_C or X_L. At low frequencies, X_L will be low and X_C will be high; for all practical purposes the circuit is an inductor. At high frequencies, X_C will be low and X_L will be high; the circuit is essentially a capacitor.



At f, theory says Z should be infinite. This can't happen in real life, in part because the math would again have to involve dividing by zero. According to the product-over-sum formula for parallel-connected components,

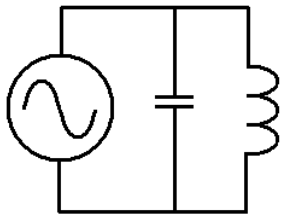
$$Z = \frac{X_C X_L}{X_C + X_L}$$

You would expect this to result in an impedance that's lower than the reactances, but remember that X_C and X_L are opposite polarity, so they add negatively – that is, they subtract. Since they're equal at f, this would leave us with zero in the denominator, which would trigger a visit from the math police. However, the real-life number in the denominator will not be zero. It is determined by parasitic resistance, which will be very small at f, meaning the impedance at f will not be infinite but will be very large.

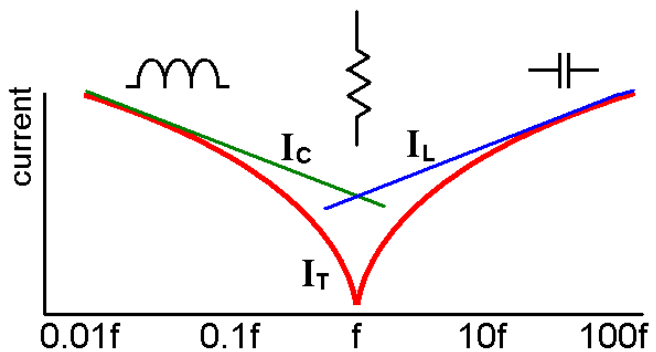
PARALLEL LC VOLTAGE AND CURRENT

In a parallel LC circuit, the voltage across each component is simply the applied AC voltage. Since we don't have a zero-Z frequency, this doesn't present any special problems.

Current flowing in L is opposite polarity from current in C, so the total current supplied by the AC source is the difference between them, $I_C - I_L$ or $I_L - I_C$. The current-vs.-frequency curve looks exactly like the impedance-vs-frequency curve of a series LC circuit.



At low frequencies, X_L is low and L conducts a lot of current. At high frequencies, X_C is low and C conducts a lot of current. But at f, where $I_L = -I_C$, current supplied by the AC source is theoretically zero.



RESONANCE

There's a story about a blind man who accidentally grabbed an elephant's tail. From then on, he would always insist that he knew exactly what an elephant is like, because he'd held one in his hand: an elephant is like a rope. And he was right, in a certain way, but in another way he didn't know what an elephant is like at all. That's the position we're in now. I've told you all about voltage, current and phase in an LC circuit, and yet haven't said a word about its most important and useful aspect:

An LC circuit rings like a bell.

There's a basic physics principle called “conservation of energy” which states, in simplest terms, that energy is never lost. When a resistor dissipates energy, the energy isn't destroyed, it's turned into heat and passed into the atmosphere. But (ideal) inductors and capacitors do not dissipate energy. So what happens to the energy in an LC circuit? At f, where $X_C = X_L$, it gets shuffled back and forth between the inductor and capacitor.

MECHANICAL RESONANCE

I'm going to approach this sideways, by looking at the mechanical equivalents of an inductor and capacitor: a mass and a spring. After all, as noted back in the first chapter of EM1, different kinds of energy are equivalent.

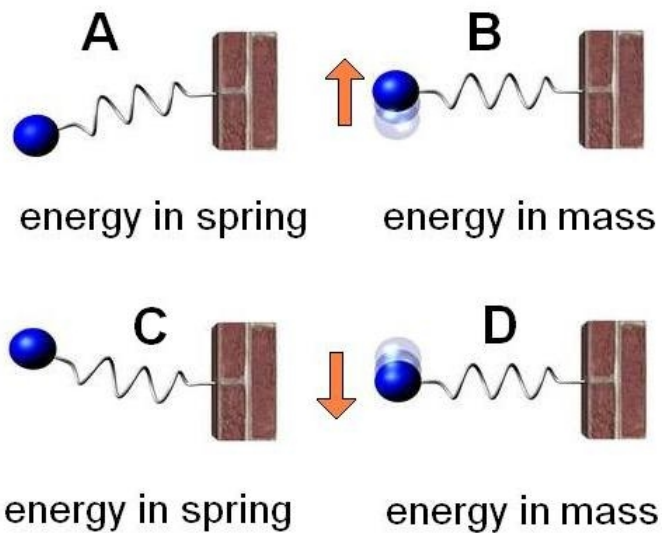
People who design mechanical systems often use electrical analogies, called **equivalent circuits**, and vice versa. Early radio pioneers called on their training in mechanics to develop electronic science. In the 1970s, an electrical equivalent circuit of a loudspeaker resulted in the first full understanding of how air behaves in a ported or vented speaker enclosure.

There are two kinds of mechanical energy. **Kinetic** energy is the energy of motion, the energy contained in a massive object that is moving. An often-used word is “momentum.” **Potential** energy is the energy of a coiled spring. Nothing is moving, but you can't say there's no energy in the system; it's just stored.

Referring to the illustration at right: If you put a mass at the end of a spring and attach the other end of the spring to something solid, and then pull the mass sideways, you put energy into the system. At first (A), all this energy will be be potential energy, stored in the bent spring. But when you let it go, as the spring pushes on the mass, it will start to move.

If we skip forward a bit, we'll see that the mass is at its original position (B), before we bent the spring. Since the spring is now relaxed, it no longer contains energy. But now the mass is moving – in fact, this is the moment when its speed is greatest. So that's where the energy has gone; it's been transformed from potential energy in the spring to kinetic energy in the mass.

Because of the momentum of the mass, it will keep going. It will start to bend the spring the other way, and the spring will start pushing back.



If we skip forward in time again (C), we will come to a moment when the spring is fully bent, and has been pushing back on the mass long enough to stop it. Since the mass is not moving at this point, it contains no energy. All the kinetic energy has been changed back to potential energy in the spring.

Now the spring is pushing the other way. Skipping ahead in time (D), we come to a moment when the spring is relaxed; all the energy has been changed to kinetic energy in the ball.

Energy would be moved back and forth this way and the oscillating motion would continue forever if it weren't for friction losses.

For every specific combination of mass and spring, there will be a specific oscillation frequency, a specific number of times per second that the mass goes back and forth. This is called the **resonant frequency** of the mass/spring combination. It's the frequency we've been calling "**f**."

If you were to graph the motion of the mass over time, you would find that it is moving in pure harmonic fashion; that is, it's making a sine wave.

You can change the resonant frequency by changing either the mass or the spring. For example, if you substitute a smaller mass, the resonant frequency will rise. The same thing will happen if you substitute a stiffer spring.

Any mechanical system that resonates involves a mass/spring combination. If you strike a chime, it rings, which means parts of the chime are moving masses and parts of it are flexing springs. The air in an organ pipe is divided into zones where air masses move back and forth and zones where air springs compress and expand.

Resonant Tuning

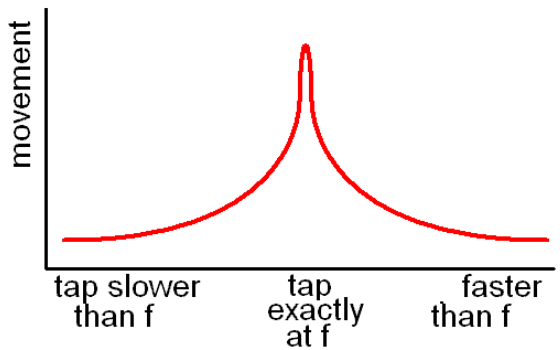
What happens if you begin with the mass-spring system at rest, and start tapping lightly on the mass? Well, it depends on how fast you tap. If you tap much faster or slower than the resonant frequency, not much happens. But if you time your taps to exactly the resonant frequency, the mass and spring will start moving. After a while they'll be moving a lot, even with the lightest taps.

Think of a grandfather clock pendulum, where gravity takes the place of the spring. On each swing, the mechanism gives a push that's about equivalent to a flea's sneeze. Yet it's enough to keep a five-pound pendulum swinging.

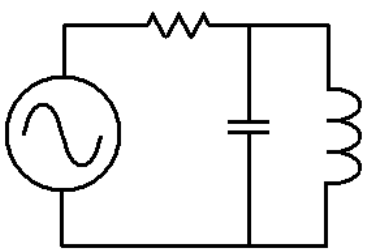
The reason a small amount of input power eventually results in large movement is that the power in a mass-spring system is not dissipated; it stays in the system, shuttling back and forth between the mass and the spring. Each tiny tap adds more energy.

The reason tapping on the mass only causes significant movement if it occurs at the resonant frequency is that the system can only contain energy, and trade it back and forth in proper resonant fashion, at that one frequency. Tapping at some other frequency may momentarily move the mass, but it doesn't input resonant energy.

The illustration at right shows how a resonant system responds to energy input from tapping on the mass. Only if you tap at exactly the resonant frequency will the system eventually build up a large movement.



RESONANT LC CIRCUITS



Since a capacitor is like a spring and an inductor is like a mass, when you put them together, the same thing happens: an LC combination will have a resonant frequency, which we call "f." The electrical equivalent of tapping a mass/spring system lightly is feeding a sine wave into an LC circuit through a high series resistance.

The formula for the resonant frequency is:

$$f = \frac{1}{2\pi\sqrt{LC}}$$

where L is in Henrys and C is in Farads. This can be simplified to:

$$f = \frac{0.159}{\sqrt{LC}}$$

For the remainder of these study guides, you are hereby given permission to round 0.159 to 0.16, which throws off the calculation by about 0.6%.

CHANGING f IN AN LC CIRCUIT

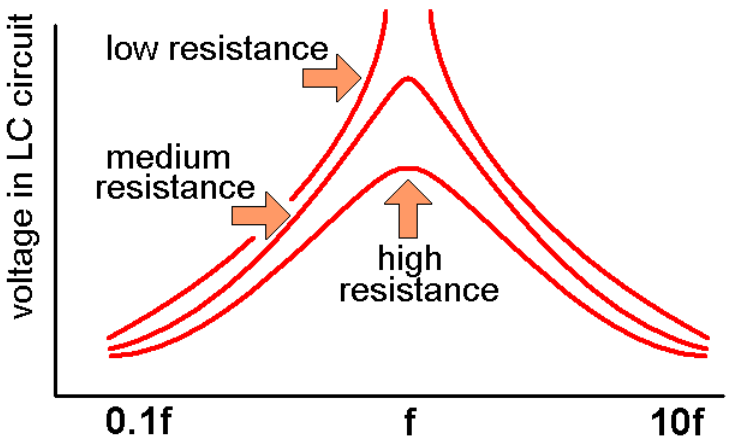
Note that f is determined by the combination of L and C. That square root sign means that if you change L or C by a certain percentage, you only get the square root of that change in frequency.

For example, if you want to change the frequency by 4:1, you have to change L or C by 4^2:1, which is 16:1.

In traditional tunable circuits, this makes it hard to tune over a wide range. One of the reasons the AM radio band covers just over a 3:1 range is that it's hard to build a variable capacitor with more than a 10:1 range.

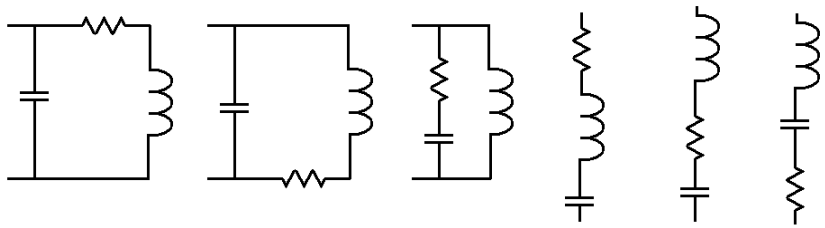
RLC CIRCUITS

Time to get real: let's start taking resistance into account. Since we're always stuck with resistance, we might as well make a virtue of it and use it as a control technique.



A high quality LC circuit only responds well to energy input at f. Deviate half a percent, and its response drops off sharply. That may be too narrow. We often want the circuit to respond to a not-so-narrow band of frequencies centered on f. We do that by deliberately screwing up the resonant behavior to a degree. This causes the response at frequencies either side of f to be attenuated even more, but the response at f is *very greatly* attenuated – thus the band of frequencies that get approximately equal treatment is broadened.

And how do we screw up the resonance? By adding resistance. If the resonant circuit were a machine, resistance would be sand in the gears. It usually doesn't matter where we put it, or if we split it up into smaller resistances.



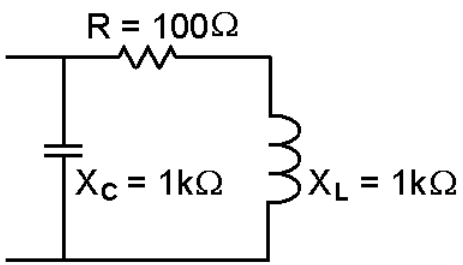
"Q"

Q stands for the resonant **quality** of an LC circuit. A high-Q circuit is one where the total reactance is much greater than the total resistance. With real-world capacitors and inductors, a Q of several hundred can often be obtained, if that is desirable. A low-Q circuit is one where resistance is high enough to constitute a sizable portion of the total impedance.

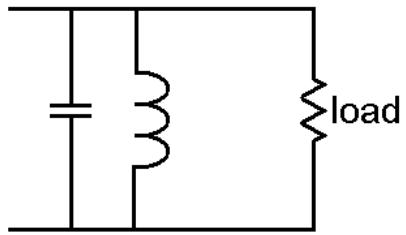
The formula for Q is:

$$Q = X / R$$

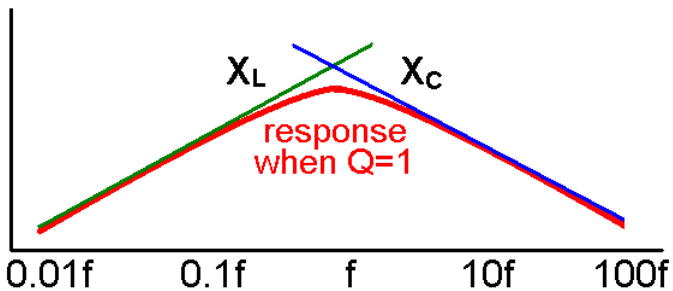
where X = X_C or X_L (they're equal at f) and R is total resistance. If we want the highest Q we can get, we use the lowest-resistance L and C we can get, and wind up with a Q of several hundred. If we want a lower Q, we add resistance. The circuit at right, if L and C have no resistance, will have a Q of 10.



The Q of an LC circuit can be greatly affected by what it's connected to. Even if there isn't much resistance in the LC circuit, a resistive load across a tank or one component of a series LC circuit will drag down the Q.



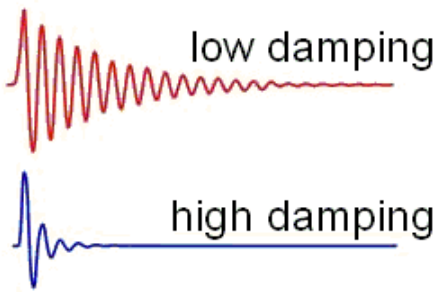
When the Q is quite low, around 1, the response curve has nearly the same slopes as L and C impedance. It rises linearly with rising frequency, doubling with every doubling of frequency, until it gets to f, then falls linearly, halving with every doubling of frequency.



DAMPING

I said earlier that an LC circuit rings like a bell – if you hit it with a voltage spike, it will resonate for a significant time, like a chime struck with a hammer.

What “a significant time” means is greatly affected by Q. If you hit a dowel rod, it may resonate at the same frequency as the chime, but the chime will ring for a whole minute, while the dowel rod will ring for maybe a tenth of a second. The English words “clang” and “clank” denote this difference.



The speed with which oscillations in a resonant system die out is called **damping**. The figure at right shows what happens when an LC circuit is whacked with an electrical transient. If damping is low, the circuit will oscillate at its resonant frequency for some time, like a chime. If damping is high, oscillation will die out quickly, like a chime wrapped in a blanket.

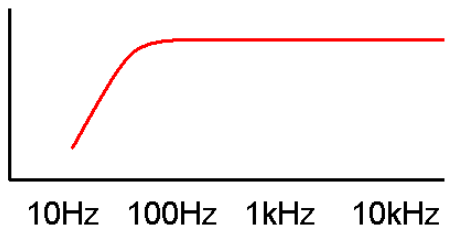
Damping is the reciprocal of Q:

$$D = 1 / Q$$

A circuit with a Q of 10 would have a damping factor of $^1/_{10}$ or 0.1.

FILTERS

A filter is a circuit that passes some frequencies and attenuates or blocks others. The range of frequencies passed with little attenuation is called the **pass band** or **passband**. The range of frequencies that is attenuated is called the **attenuation band**, or if the attenuation is very great, the **stop band** or **stopband**.

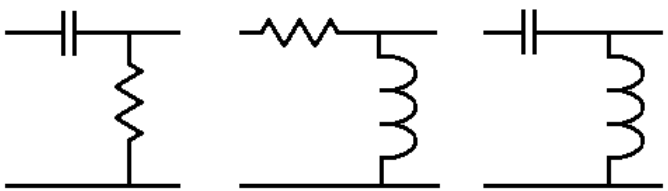


A circuit's behavior with respect to frequency is called its **frequency response**. The curve at right, for example, is the response of a “blast filter,” which attenuates low frequencies to remove breath, pop, handling and wind noise from a microphone signal.

The frequency response of a filter is generally calibrated in decibels. We won't get to those until EM3; thus, we can't examine responses in great detail here. However, we will look at the names and functions of the most common types so you'll know them when you see them. Most filters can be built in RC, RL or RLC form. They can also be implemented as “active” filters, incorporating amplifiers to modify their characteristics, or in digital form. We will also study those circuits in EM3.

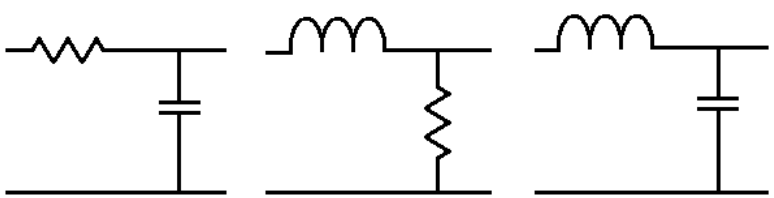
High-Pass

A high pass filter passes signals above a chosen frequency, called the **cutoff** or **turnover** frequency, and attenuates those below. RC, RL and LC high-pass filters are shown below.



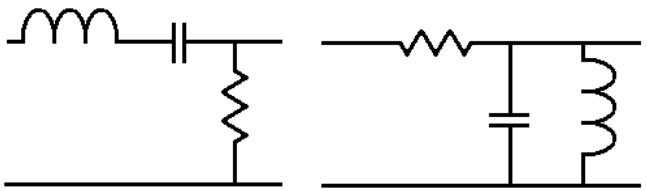
Low-Pass

The low pass filter is the mirror opposite of a high pass filter; it passes signals below the cutoff frequency and attenuates those above it, as shown on the next page.



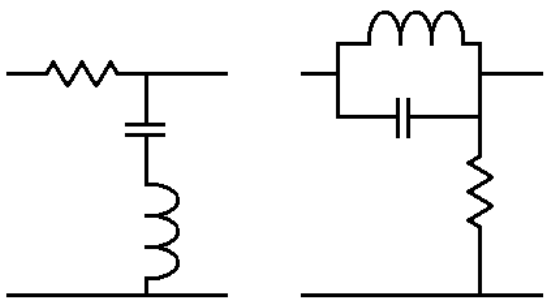
Band-Pass

A band-pass filter passes a selected band of frequencies, attenuating those above and below. A wide band pass is obtained by connecting a high-pass and a low-pass filter in series. A narrow bandpass is obtained by using a tuned LC or RLC circuit. Two RLC variations are shown below.



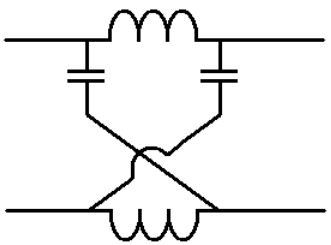
Band-Stop or Band-reject

A band-reject filter attenuates a selected range of frequencies, passing those above and below. The frequency band rejected is usually narrow, to **notch** out a single frequency that's causing problems.



A band-stop filter is also called a **trap**, as in trapping an offending signal, or a **notch filter**. Suppose you live next to an FM station, which comes blasting through your radio no matter what station you're trying to tune to. A trap inserted in the antenna line and tuned to that station's frequency would allow you to listen to other stations.

Band-reject filters are typically RLC circuits, because only a resonant circuit can achieve the narrow notch usually required. At lower frequencies, RC circuits called “twin-T” and “bridged-T” filters can be used. We will study those in EM3.

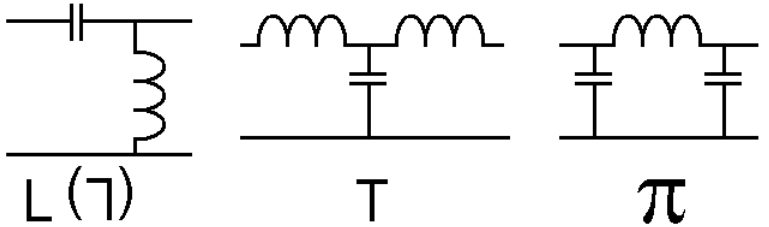


All-Pass

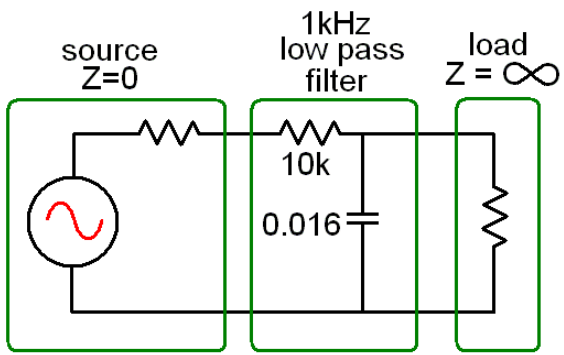
An all-pass filter does not attenuate any frequency. Its purpose is to change the phase of some frequencies relative to other frequencies.

L, T, π FILTER DESIGN

These designations refer to the parts layout of a filter. The letters describe their shape, inverted in the case of an L-filter. T and π filters are excellent filters with very high attenuation in the stopband. They're extensively used in RF work.

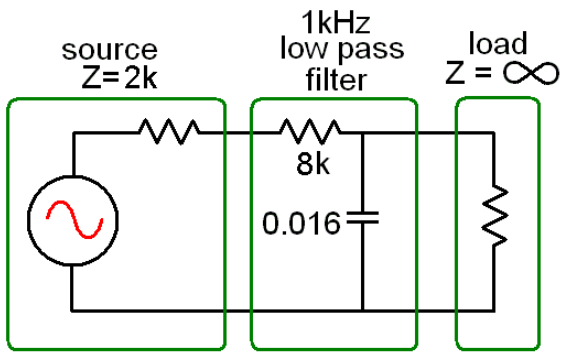


SOURCE AND LOAD IMPEDANCES

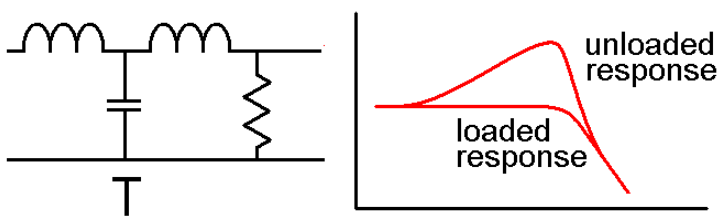


At the beginning of this chapter, I said we'd assume zero source impedance and infinite load impedance. That's handy for analysis, because then our math doesn't have to take into account anything but the R, L and C components of the filter. A great many filters do operate in that environment, especially when they're found inside equipment.

When the source or load deviates from this ideal, we can change a value in the filter to compensate. Here, for example, the source has 2kΩ of resistance, so R in the filter has been reduced by 2kΩ.



RLC filters usually *must* be loaded. A lowpass T-filter, for example, goes all haywire without a load.

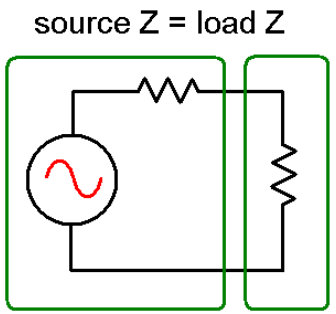


Calculating response when source or load are different from the ideal is often complicated, but the tech seldom has to do it. The main thing you have to watch out for is that an existing filter that you run into on the job is connected to the source and load impedances it was designed for, which will either be labeled or implied by the application.

MATCHED-IMPEDANCE FILTERS

In RF and digital data distribution, the **matched impedance** convention is almost always followed, where the impedance of a signal source is the same as the impedance of its load. Usually, a single impedance will apply everywhere in a system. For example, in an RFsystem every antenna and every equipment input and output is typically 50Ω, except in video systems where the standard impedance is 70Ω. This is explained in much greater detail in chapter 10.

When a filter is used in a matched system, it must be designed to work with the system's source and load impedances. Also, the filter's own input and output impedances should be the same as the system impedance. This is often difficult to achieve, but it's often satisfactory if the filter just matches the rest of the system over the range of its passband.



There is much more to say about filters, and you'll find it in EM3.

End of Chapter 9

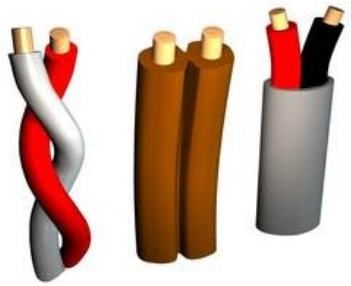
EM2 Chapter 10
CABLING

Objectives of this chapter:

- define “cable”
- explain impedance and its causes, and why it is often necessary to maintain impedance in a system
- explain “propagation” and its relationship to cables
- define “termination” and explain the effects of proper and improper termination
- explain grounding conventions in electrical and electronic work
- discuss shielding, its purpose and limitations
- discuss coaxial cable
- discuss balanced and unbalanced cabling, twisted pairs
- explain how to prepare shielded cable ends for connection

“CABLE (KAY-bul) n. 1. A potential source of trouble connecting two other potential sources of trouble.”
- anonymous

CABLES



Okay, here's the real definition: two or more wires bundled together and intended to be treated as a single unit are referred to collectively as a **cable**. The wires may be held together by twisting them around each other, by physically attaching them to each other as in lamp cord, or by enclosing them in a flexible rubber or plastic tube called a **jacket**. There may be two wires or two hundred.

THE INTERCONNECTION PROBLEM

Electronic communication of any kind – audio, video, data, whatever – usually involves one piece of equipment handing off a signal to another piece, which may hand it off to a third, like the baton in a relay race. Signal may be sent at least three ways:

- optically, over a light beam, which we'll study in EM3.
- via electromagnetic radiation, which we'll study in EM6.
- over wires, which we'll study in this chapter.

Getting a signal from point A to point B is easy; the hard part is getting it there without ruining it. A signal *always* gets damaged as it passes through electronic circuitry; the transfer is never perfect. It is the job of the engineers who design electronic gizmos, and the techs who interconnect them, to see to it that the damage is minimized. So, we're going to spend most of this chapter looking at how to avoid the many potential pitfalls.

When the interconnection between two pieces of equipment is via cable, signal degradation comes primarily from four causes:

- The cable itself may degrade the signal.
- Signal may be damaged by an improper interface at one or both ends.
- Interference may be picked up from external sources.
- A signal may be measured against the wrong reference point.

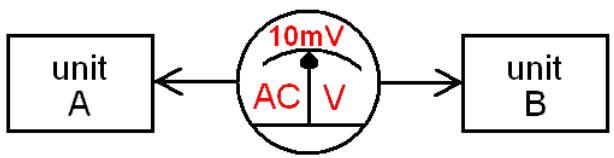
We're going to take these in reverse order and study references first. Since we're going to be talking about interconnections between pieces of equipment, and “pieces of equipment” is cumbersome, I'm going to refer to them as **units**.

GROUNDING

Grounding is the science, and to some extent the art, of establishing common reference points. At the risk of boring you by repeating this yet again:

- A voltage is a difference
- that is, it's a potential
- that is, it's measured against a reference, which most often is called **ground**.

So what? Here's what: in a complex system, there are a number of grounds. Just because they're all grounds, that doesn't mean they're all at exactly the same potential, DC or AC. Two grounds may have a potential of 10mV between them, or 100mV, or they may have no relationship to each other at all.

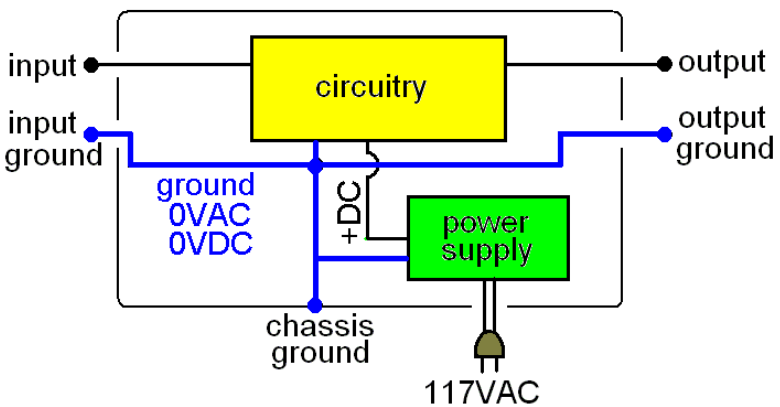


If a signal is referenced to the wrong ground, whatever voltage difference exists between the grounds will become part of the signal. That's degradation. So we have to pay as much attention to grounding as we do the signal.

Independent Grounds

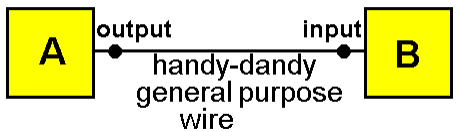
As you know, the Earth is spinning. It's also revolving around the Sun, which is revolving around the center of the Milky Way, which is moving through space. But space isn't our reference; the Earth is. As far as our senses can tell, when we stand still, the ground we're standing on is stationary.

Likewise, the frame of reference for the circuitry inside a unit is that unit's own *internal* ground. Within a unit, there is a central ground that every voltage is measured against. The input reference, output reference, circuitry reference, power supply reference and chassis are all connected to this ground, which is defined as zero volts.



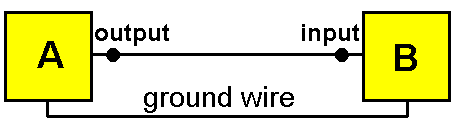
The unit's ground may have 500V on it, with respect to Earth. The unit doesn't know. All it knows is what the voltages at its input, output, and at various points in its innards are, *compared to its own ground*.

Now suppose we have two units, “A” and “B.” Unit A is putting out 1 volt, and we want to connect that to unit B's input. If we just take a wire and connect output A to input B, will input B “see” 1V? Not likely. The voltage will be completely random, because unit A is putting out 1V referenced to *its* ground, but unit B is measuring the input voltage against *its* ground, and the two grounds have no specific relationship to each other.



If we want to do what we set out to do, which is get a signal from A to B, we have to run another wire and connect the two units' grounds together, so they're both looking at the same reference.

Okay, no big deal. You can run two wires about as easily as one. Connect one wire from output to input, connect the other from ground to ground. Problem solved, right?



For the time being, I'm going to say yes. The primary requirement has been met; units A and B are singing off the same sheet, so to speak. When A says, “I'm putting out a signal,” and B says, “I'm seeing a signal,” they're both measuring against the same ground reference.

In practice, it's not that simple. “Properly” connected grounds can still permit signal degradation, or even cause it. But that discussion will make more sense if we set it aside for the time being, study some other things, and then come back to it.

NOISE

In chapter 3, we said the word **noise** has two definitions. The one we studied in chapter 3 was white noise generated by the movement of atoms and molecules. The other definition of noise is **contamination of a signal**, which we're going to deal with now.

A real-world signal can be thought of as a combination of perfect signal, whatever “perfect” may mean in a given case, and some degree of contamination. For example, nobody has ever generated a perfect sine wave; the best we can do is something close to a sine wave. These days, we're routinely getting down to 0.003% harmonic content, which is very close indeed. We can view this as the sum of a 100% perfect sine wave plus 0.003% of grunge.

The contamination can be white noise generated by molecular motion, which no system is entirely free of. But it can also be 60Hz hum, lamp dimmer buzz, an unwanted “sound track” from the nearest radio station, transients from a furnace igniter, cross-coupling from some other signal, and a thousand other things. No matter what the contamination consists of, if it isn't supposed to be part of the signal, it is defined as noise.

The ratio between the original signal and the noise is called the **signal-to-noise ratio**, abbreviated **S/N ratio** or just **S/N**. It may be expressed as a percentage but is more commonly expressed as a **decibel** or **dB** ratio. Our study of decibels is coming in EM3, but just for reference, here are some very rough benchmarks (next page):

Noise voltage as a percentage of signal voltage	S/N	example
10%	20dB	noisy cafe, hard to talk
1%	40dB	office
0.1%	60dB	medium grade mp3
0.01%	80dB	decent CD

These percentages translate to voltage ratios: if a 1-volt sine wave has 10mV of crud on it, that's 1% crud and the S/N ratio is 40dB. Although the above examples are sounds, the principle also applies to video, data, or any kind of signal at any frequency or combination of frequencies.

The most important thing to understand about noise is that, once it becomes part of the signal, it is difficult or impossible to separate out again. That's why we must take steps, at every point in the signal chain, to prevent noise from getting into the signal in the first place.

The two main ways noise can get into a signal are:

- radiation and induction, via air and space
- conduction, via incorrect grounds

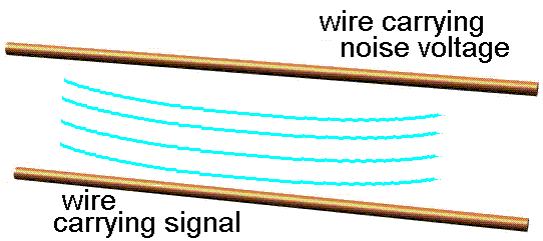
RADIATED AND INDUCED NOISE

If there were no such thing as noise, there would be no reason for cable to exist other than the convenience of a neat package. People would just string a signal wire and a ground wire between units any way they liked, and signal would get from point A to point B in good shape.

But in the real world, there is a lot of radiated crud floating around. Even if we weren't making it, there would be noise from the static electricity associated with clouds and storms.

Radiated noise usually can't get at the signal while it's confined inside a unit, but when a signal moves out into a cable, it's venturing into a hostile world. If we don't take special precautions, the signal will pick up enough noise by the time it reaches its destination to make it useless.

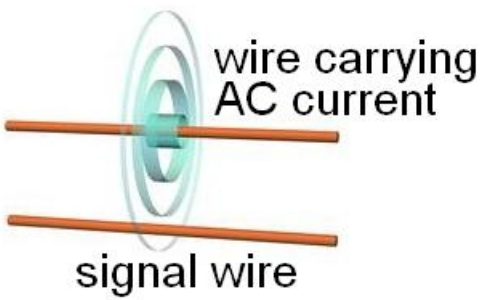
Electrostatic and Electromagnetic Coupling



There are two ways noise can be wirelessly coupled from a noise source to a signal wire. One is **electrostatic** coupling. If we think of the noise source as one plate of a capacitor and the signal wire as the other, we can see how interference can be transferred in the manner of a coupling capacitor.

It isn't a very good capacitor, but even a bad capacitor can cause trouble if the interference source is much higher voltage than the signal, as is often the case. When the interference source is a 600 volt AC circuit powering machinery and the signal wire is carrying a 6mV signal, 0.0001% coupling will result in a lousy 20dB S/N ratio.

The other kind of coupling is **electromagnetic** coupling. AC power transformers, fluorescent lights and wires carrying high current all generate 60Hz magnetic fields, and these can be inductively coupled into other wires.



A wire can also pick up energy from nearby wires carrying other signals. This is called **crosstalk** and it can be electrostatic or electromagnetic or both. If you've ever been on a telephone and heard faint, distorted voices from other people's conversations, you've heard crosstalk. Though a signal is originally “signal,” when it gets into another signal channel where it doesn't belong, it's noise.

How do we combat radiated noise? There are four principal methods:

- high signal levels
- distance and orientation
- shielded cables
- balanced wiring

Let's look at each of those.

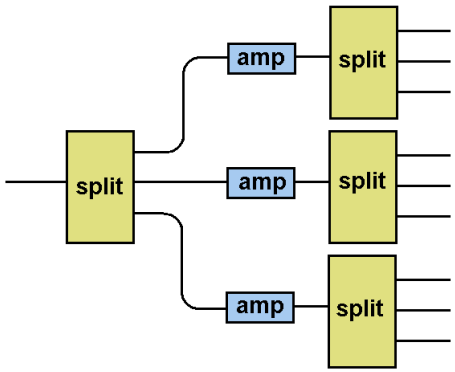
SIGNAL LEVELS

We can combat noise by sending the signal at such a high level – techs call this a **hot** signal – that noise is less significant by comparison. If we can increase the level of a signal coming out of a unit by 10dB, that's as good as decreasing noise by 10dB. If we're talking about sound, something that is 10dB lower in level than it used to be is subjectively perceived as being half as loud. Cutting noise in half is worth doing.

One case where a hot signal is helpful is a distributed video system like the ones installed in apartment buildings and large homes. By the time the signal gets split out to all the different screens, the level may have dropped so low that the S/N ratio is marginal. If an amplifier can be inserted early in the signal chain, this can be prevented.

But amplifiers have limits. The hotter the signal, the more expensive the circuitry to handle it. That's one reason why consumer equipment uses lower signal levels than professional equipment.

Amplifiers can overload. In a big video system, for example, installers often find that they can't put a single amplifier at the head end. They have to split the signal, then insert amplifiers, then split again.

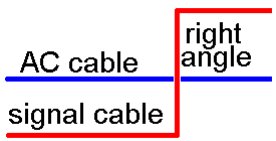


Also, a hot signal may become a source of interference itself. Telephone companies set strict limits on the signal levels that can be sent over their land lines to prevent crosstalk. And in some cases – a microphone, for instance, or a receiving antenna – the signal sees a cable before it ever sees an amplifier. So, using hot signals is a good technique, but it doesn't solve the whole problem.

DISTANCE AND ORIENTATION

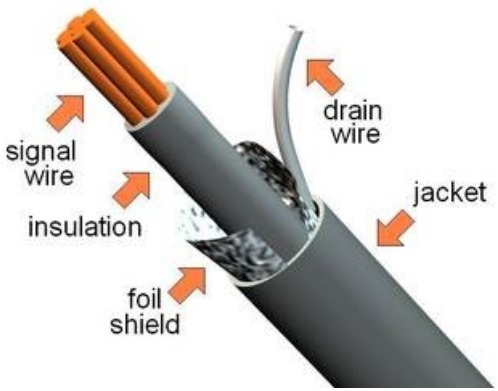
The simplest way of reducing noise is to run a signal cable as far away from the noise source as possible. Electrostatic and especially electromagnetic coupling falls off very quickly as distance increases. This means you never attach a cable carrying a sensitive signal to an AC power conduit. (In addition, it's illegal to do that in most jurisdictions.) In an equipment rack, find out where the power transformer in each unit is and make sure the signal wiring doesn't go near it. This is not as important if all the signals are RF or data, because the frequencies are so high that the associated equipment is probably “blind” to 60Hz noise.

If your signal line must cross an AC circuit, cross it at right angles. This minimizes the area of the signal wire that's near the AC wire, and as you probably know a magnetic field running along a wire, rather than across it, does not induce a voltage. So a right angle minimizes both electrostatic and electromagnetic coupling.



SHIELDED CABLES

Noise can be reduced by enclosing the signal wire in a protected environment using **shielded** cables. In a shielded cable, the ground wire is made into a tube that completely surrounds the signal wire. This prevents electrostatic interference from getting at the signal. A cable made this way is called a **coaxial cable** or **coax**, because both conductors are on the same axis instead of side by side.



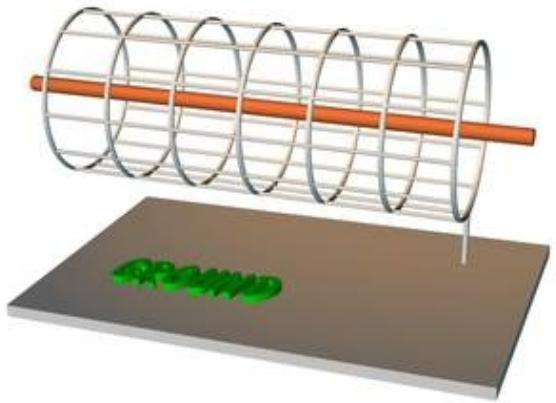
The illustration at left shows one kind of shielded wire. The shield enclosing the signal wire is aluminum foil or aluminum-coated plastic. A **drain wire** contacts this shield all along its length. The drain wire provides a way to electrically connect to the shield; much more convenient than trying to make a connection to foil.

There is a lot of confusion about the function of a shield. Many people think of it as a brick wall that blocks interference. But a shield is nothing more than a fat wire, and it can pick up interference and couple it into the signal wire.

not how a shield works

The true function of a shield comes from the fact that it is attached to the circuit grounds of the units it is connected to. Being electrically grounded, and a conductor, it encloses the signal wire in a ground-potential environment. Incoming interference is intercepted and drained off before it can reach the signal wire.

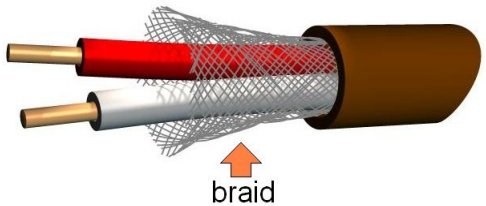
The point about grounding is crucial. A shielded wire where the shield is left unconnected is no better, and often worse, than an unshielded wire. *An ungrounded shield is an antenna*.



When I say “ground” I mean *unit ground, not Earth ground*. This is one of the hardest things for people to grasp: in most cases, the presence or absence of an Earth ground is irrelevant. An airplane has lots of shielded wires in it, and there is certainly no possibility of an Earth ground at 35,000 feet – yet the equipment works.

Remember what a signal is: a voltage measured against the circuit grounds of the units it is connected to. We want to enclose the signal wire in an environment that is zero volts *with respect to the units connected to it*. That means connecting to their circuit grounds. Remember, too, that the shield is also serving as the zero-volt ground reference for the signal.

Braided Shields



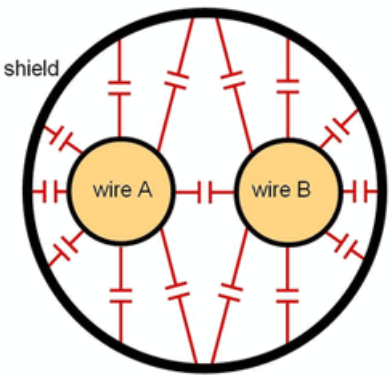
Where physical flexibility is needed, the shield usually takes the form of a copper or aluminum **braid**. The tighter the braid, the better the shielding. Sometimes a cable with a loose braid is used as an RF antenna where the purpose is to let little bits of signal leak out all along its length. Drive-in movie theaters use this technique a lot.

RF and video cables often use a braid over aluminum foil. The braid itself may be aluminum. This wouldn't work for cables where the connectors are soldered on, but with RF and video cables the connectors are usually crimped on.

Some cables have two or more separate layers of shielding, on the principle that you can do a better job of knocking interference down if you take several whacks at it.

Shielding Limitations

The main limitation of cable shielding is that it only stops electrostatic interference. Let me say that again, because it's such a common misconception – I would bet 80% of professional technicians don't understand this. Cable shielding only stops *electrostatic* interference. Against electromagnetic interference, it's completely useless. I have to laugh when people insist on using shielded cable to prevent interference from fluorescent lamp ballasts. The field generated by a ballast is entirely magnetic. If there's really a problem, shielded cable won't fix it.



Shielded cable has higher capacitance than unshielded cable, so its losses at high frequencies are greater. Moreover, connecting a highly capacitive load to a stiff amplifier can cause it to become unstable. For that reason, shielded cable should *never* be used for speaker wiring.

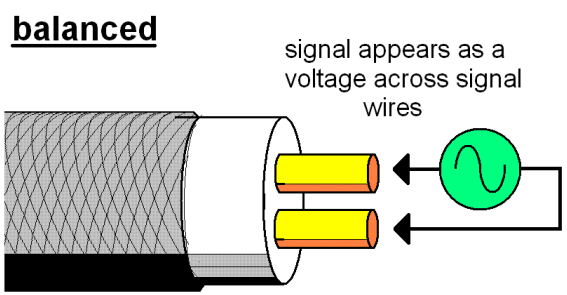
Shielded cable's other limitation is that a grounded shield is only really grounded up to a certain distance from whatever it's attached to. You recall from EM2 that the frequency of an electrical waveform corresponds to a wavelength. At a distance of about ¼ wave from the shield ground point, it ceases to be effective. At audio frequencies, that is measured in miles. At 100MHz, it's a couple of feet. At high radio frequencies, there is no such thing as a ground at any significant distance from the unit that you're grounding to.

This doesn't make the shield useless; at high frequencies the cable becomes a sort of tube that guides the waves along its length. But don't expect it to do the grounding job it would do at low frequencies.

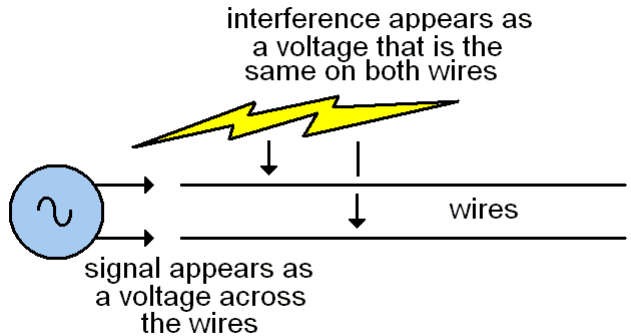
BALANCED WIRING

The fourth and by far the most effective way to prevent interference pickup is to use **balanced** cabling. In a balanced circuit, the desired signal appears in a different form than undesired noise, and this allows the receiving unit to tell them apart.

A balanced cable has two identical signal wires. The signal appears as an AC voltage *across* the two wires. In a balanced circuit, whenever there's a positive voltage on one wire, there's an equal negative voltage on the other. This is called a **differential-mode** signal.



The voltage induced by radiated noise, on the other hand, is likely to be the same on both wires. This is a **common-mode** signal.

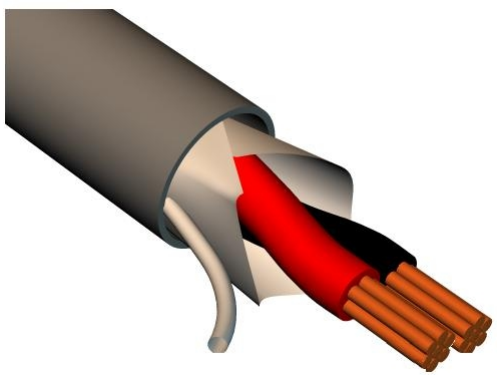


The magic of a balanced circuit is that it forces noise to exist in a different form from the signal. If the receiving unit is designed to be sensitive only to a differential-mode signal, if it can only "see" a voltage difference across the two wires, then it will be "blind" to the common-mode noise.

The critical thing is that the noise voltage must be exactly equal on both wires. The two wires should run right next to each other, and ideally should be twisted around each other. That way it will be almost impossible for a noise source to affect one wire differently from the other. You'll hear the expression **twisted pair** a lot. It refers to a pair of wires twisted around each other and connected into a balanced circuit.

Going the other way, a balanced, twisted pair also radiates less signal outward. Between reduced radiation and reduced pickup, this essentially eliminates the problem of crosstalk between pairs, especially if the pairs in a cable all have different numbers of twists per hundred feet. Cable manufacturers know this and make their cables that way.

In many cases, balanced wiring eliminates the need for shielding, which is a very good thing. A CAT5 data cable has four balanced pairs. A telephone exchange has thousands. Can you imagine the mess and bulk if each of those pairs had to be individually shielded?

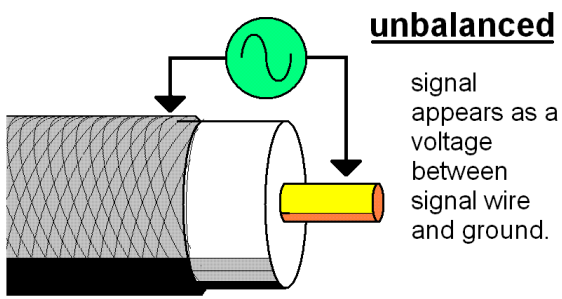


However, for very weak signals, a balanced line plus shielding gives the best protection. As a rule, microphones use balanced, shielded cables. Sensitive laboratory equipment often uses balanced, shielded pairs. Recording and broadcast studios are cabled with balanced, shielded cables, not so much because shielding is useful in a balanced circuit but because things sometimes accidentally become unbalanced.

The effectiveness of balancing is shown by the fact that, in the days when the telephone system consisted of land lines, a long-distance call could go across the USA on an unshielded pair, bundled with thousands of other pairs, and still arrive at the other end with minimal crosstalk.

Non-Balanced Circuitry

Balanced circuits are a special case. A signal usually appears on one wire and is referenced to circuit ground. The internal circuitry of equipment is almost always non-balanced; otherwise, the manufacturer would have to provide two of everything in a place where interference pickup isn't a problem. A non-balanced signal is called **unbalanced**, an unfortunate term because it implies that the signal is defective, which there's no reason to assume.



Most signals are unbalanced. Another name for this is a **single-ended** signal.

Balanced/Unbalanced Conversion

When the unbalanced circuitry inside a unit is connected to a balanced cable, a conversion must be made. The three most common ways of doing this are:

- active balancing/unbalancing, which involves special circuits that perform the conversion at a unit's input or output
- pseudo-balancing, which can be used in some cases, involving a resistor to simulate half of a balanced circuit
- transformer conversion

The first two methods involve circuitry you won't study until EM3, so we're going to ignore them for the moment. Transformer conversion is the oldest and simplest method, and in some ways still the best.

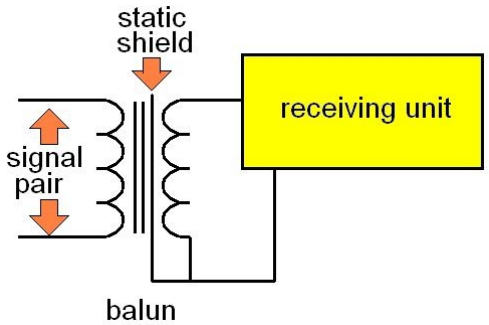
A transformer used to convert an unbalanced circuit to balanced or vice versa is called a **balun**, short for **balanced/unbalanced**. A transformer is the natural, obvious interface device for a balanced line, because the signal on either side of the transformer appears as a potential across a winding – in other words, a differential signal. Only a voltage *across* the primary will induce a voltage in the secondary. So, by nature, the transformer is exactly what we want: a device that is blind to a common-mode voltage.

If the primary is connected only to the signal wires, it is called a **floating** winding. This provides another benefit: electrical isolation. When the signal is coming from a telephone cable miles long, and there's a thunderstorm blowing through, the wires may have hundreds of common-mode volts on them. Better not to connect that to any point in the local circuit, including ground.

We get still another benefit: we don't have to bother grounding the units on either end together. In a transformer-isolated balanced circuit, each wire is the other's reference, so a common ground is unnecessary. When the two ends of the circuit are miles apart, it's impossible to establish a good common ground. In such cases, a transformer-isolated balanced circuit is often the only thing that will work.

Just for insurance, many transformers – the ones used in modems, for example – have a copper foil shield between primary and secondary windings, to prevent common-mode noise from electrostatically coupling across the transformer. This shield must, of course, be connected to the local unit's ground.

So: we connect the transformer's primary to the balanced line and let it float. On the secondary side, one end of the winding gets connected to unit ground; the other end provides signal to the unit's input. The static shield also gets connected to the receiving unit's ground.



Transformer conversion is the best conversion method from the standpoint of maximum common-mode rejection. It's also the bulkiest and most expensive, and results in small distortions of the signal waveform. In audio, the other methods of balanced-to-unbalanced conversion mentioned have largely replaced transformers. However, there are still situations in audio where we can't get away from them.

In video, RF and data (modems), baluns are much smaller and less expensive, and they're the standard method of converting a balanced signal to unbalanced, or vice versa.

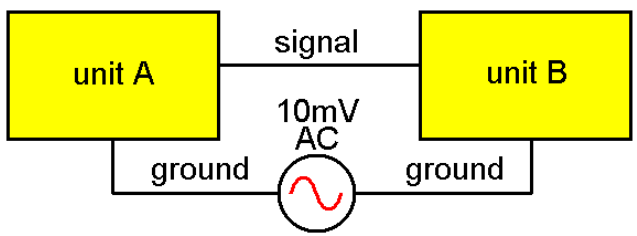
Unbalanced, Twisted Pair

Voldemort and Bellatrix. Okay, I'm sorry.

GROUNDING AND THE REFERENCE PROBLEM

Way up there about nine pages ago, I said connecting the grounds of two units together solves the reference problem – but not quite. You'd think that if you connect the metal frames of two units together with a wire, they'd be at the same voltage. Approximately, yes. Precisely, maybe not. And when you're dealing with a weak signal, or a signal that must be kept in perfect condition, that small difference becomes a big problem.

Scenario: a sound system in a church. Someone is speaking; between words and sentences there is supposed to be silence. The signal level coming out of unit A is 1V, and we want that 1V to get to unit B without degradation. The grounds are wired together, but there's something funny going on: there's 10mV of 60Hz AC between the ends of the ground wire.



This means our 1V signal, when it gets to the input of unit B, will be 1V of signal plus 10mV of hum. If the connection is not balanced, unit B will treat the hum as if it's supposed to be part of the signal.

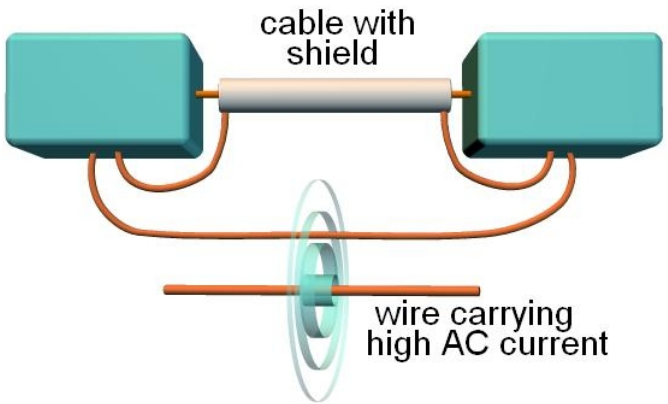
10mV of 60Hz hum isn't much, it's only 1% of the signal, but that works out to a 40dB S/N ratio, which doesn't even come close to acceptable performance in professional work. In the silences, the hum will be very clearly audible.

With video, a 60Hz ground potential can cause wide horizontal **hum bars** across the picture. With RF and data, it can degrade or disrupt communications.

But how can the voltages at two ends of a ground wire be different? Two ways: a high current in the ground circuit can cause a voltage drop, and an alternating magnetic field can induce a voltage.

GROUND LOOPS

The ground-potential problem shows up most often when there are two or more separate ground paths. This is called a **ground loop**. It's not always possible to avoid them. The best you can do is to assign a little piece of your brain to watching out for the problems I'm about to describe, avoiding them when you can, and knowing where to start looking when an issue crops up.



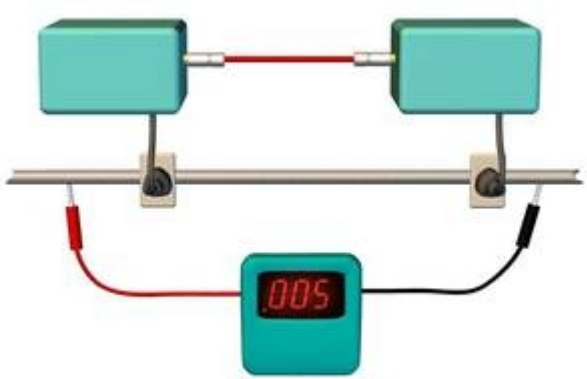
In the illustration above, the blue units are connected by two ground paths. One is the shield of the signal cable. The other is a second path, which could be an AC safety ground, the metal rack both units are mounted in, or some other conductor that's common to both. A nearby AC wire is inducing a voltage between the ends of the second ground.

The cable shield and the second ground conductor form a loop. This can be viewed as a shorted transformer winding; the voltage induced in the second ground conductor will create a high current in the loop, which will create a voltage drop across the ends of the cable shield. After all, its resistance may be low, but it's not zero.

Since there is no comparable current in the signal wire inside the shield, it will not have the voltage drop. Thus there will be a 60Hz voltage between the signal wire and the shield, which the receiving unit will misinterpret as signal.

AC Safety Ground Loops

One of the most common sources of loops is the safety grounding of electrical equipment. This is one of those cases where a good Earth ground does nothing but cause trouble. Unfortunately, for safety reasons, you can't avoid it. The rule of thumb is, if a piece of equipment has a 3-pin AC plug, the ground pin *must* be grounded to the third conductor of the outlet, unless the building's wiring is old enough to not have grounded outlets.

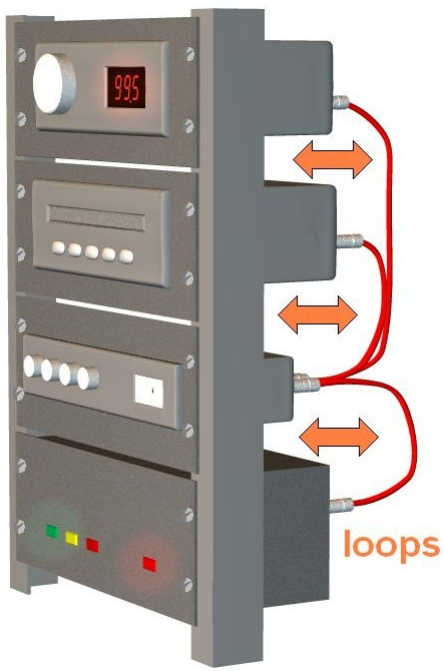


The illustration above shows a 5mV AC potential across the conduit between two outlets. This could be caused by magnetic induction, or it could be caused by a faulty piece of equipment that's putting a lot of current into the ground circuit. In large industrial buildings especially, that's more common than we'd like to think.

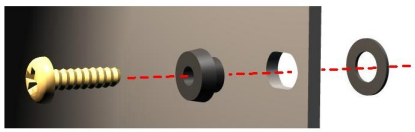
This particular problem can be largely avoided by making sure every unit in a system is plugged into the same outlet strip. Where you're receiving a signal from or sending a signal to a unit located far enough away that you cannot plug it into the locally grounded outlet, it may be necessary to insert an isolation transformer in the signal circuit to break the ground loop.

Rack Ground Loops

When a number of units are mounted in a rack, the rack becomes a ground circuit. This creates ground loops between signal cables and the rack, and between grounded power cables and the rack.



Loop problems can be reduced by bundling signal cables together and placing them as close to the rack as possible. If some units have grounded AC plugs and some don't, group the ones that do at one end of the rack.

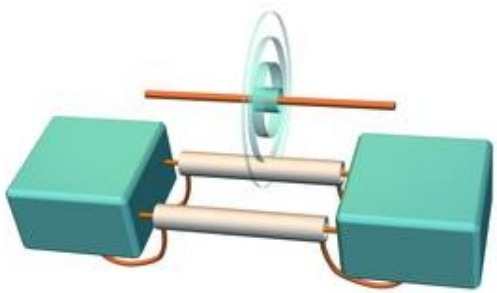


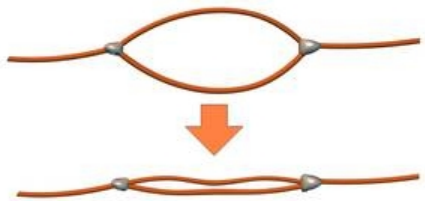
It may be necessary to mount some units using plastic shoulder washers to isolate them from the rack.

Occasionally, noise may be reduced by *not* grounding a cable shield at both ends, if there is another ground path running right alongside the cable. In this case, the other ground serves as the reference and the shield just drains off electrostatic interference. This works best at audio and low RF frequencies.

Multiple-Cable Ground Loops

If the units are connected together by more than one cable, the cable shields can form loops. These kinds of loops can be rendered harmless by bundling the cables together. Any time you can close a loop, do it. A voltage may still be induced in the ground wires, but a voltage *difference* between wires will not be induced, so there will be little ground current. This is related to the principle of twisting wires in a balanced circuit. Component video cables and stereo audio cables come bundled together for this reason.





FINDING AND FIXING GROUND LOOPS

1. Isolate the problem. In a complex system with many parts, try disconnecting one cable at a time and noting the results. You may want to start in the middle of the chain; this will quickly tell you which half of the system the problem is in.
2. Temporarily lift AC grounds. Plug grounded equipment in via a 3-prong adapter, so the AC ground circuit is broken. If the noise goes away, you may need to isolate that unit's mounting (shoulder washers) or change the wiring.

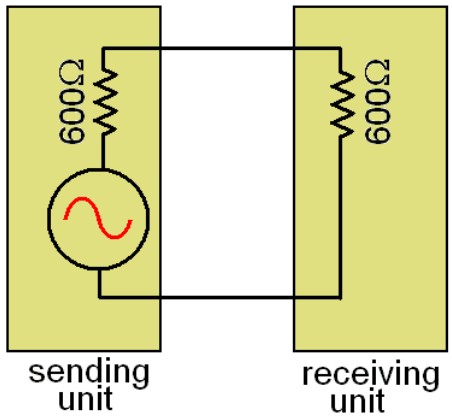
You will be tempted to just leave the 3-prong adapter in the circuit, but don't do it. Even if nobody gets hurt, you open yourself to fines and lawsuits.

3. Take a stout wire, like #12, and temporarily short the chassis of two units together. If it changes the amount of noise, the loop that's causing the problem is probably between those two units.
4. Insert a transformer in the signal line to break the chain.
5. Convert to balanced wiring if possible. This may not be an option, especially with video. The video cabling standard (75Ω, unbalanced coax) was set many years ago, when the circuitry to accommodate balanced wiring would have been a major technical challenge. If the problem is 60Hz AC, you may have to just insert a filter. Not as good as fixing the ground loop, but it does work. Such filters are available pre-made.

We've reached the point where you know about various sources of noise, how they get into a cable, and how to guard against them. Now it's time to look at some other issues you have to handle in order to get signals from point A to point B without problems.

IMPEDANCE MATCHING

There's a long-established tradition that the input impedance of each unit must be the same as the output impedance of the previous unit, so the two interfaces work together as a system. This is called **impedance matching**.

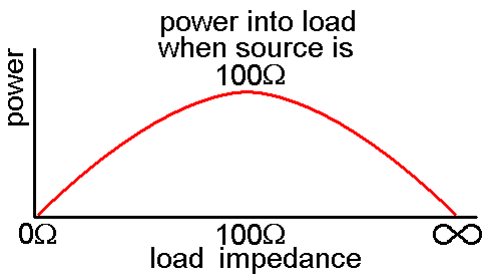


Many engineers and techs have an almost superstitious belief that source and load impedances must always be matched – superstitious because they believe it is true, but don't fully understand why. There are three reasons for matching impedances, two of which are largely obsolete.

You still have to match impedances in most cases, but I want you to know why you're doing it, and when you shouldn't.

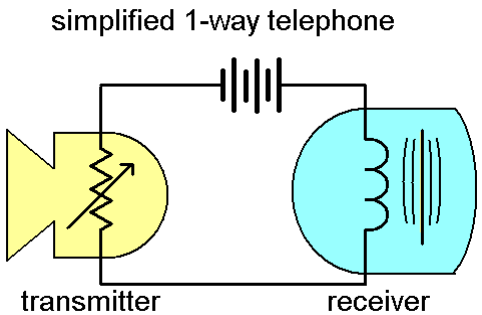
Maximum Power Transfer

Engineers have long known that, if an AC signal source has a certain internal impedance, the way to get the most power into the load is to make the load the same impedance. Suppose an AC voltage source operates into an open circuit. Since the load's impedance is infinite, no current can flow; thus no power can be transferred.



Now suppose the voltage source operates into a short circuit. A lot of current flows, but no voltage can develop across the load. All the voltage is dropped across the source's internal impedance. So, once again, no power goes to the load.

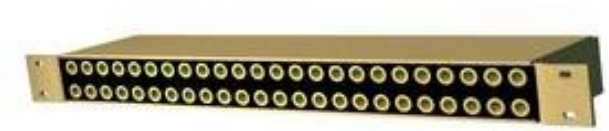
Between a short circuit and an open circuit, varying amounts of power are transferred. High impedances result in high voltage but low current; low impedances result in high current but low voltage. But there is an optimum impedance where power delivered to the load is maximized, and this turns out to be when source and load impedances are equal, or **matched**. The power-transfer efficiency in this condition is 50%, which is to say half the generated power is burned up in the source impedance and half gets to the load. If you view source impedance and load impedance as a 2:1 voltage divider, this will be obvious.



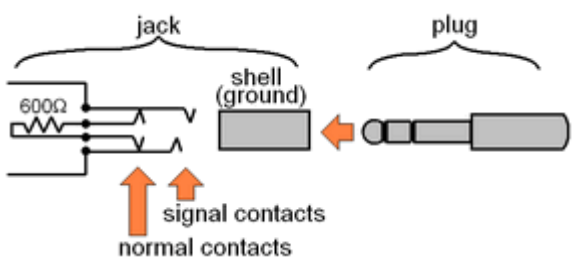
The benefit of matching impedances was established in the early telephone days. Techs call the telephone mouthpiece a **transmitter** and the earpiece a **receiver**. For over a century, the transmitter was a capsule containing carbon granules that formed a resistor. Sound waves cause momentary changes in the transmitter's resistance. A DC voltage is applied to the transmitter, causing a current to flow; the resistance changes cause current to change. The result is an AC signal with a power level somewhere in the neighborhood of 1mW. That isn't a lot of power. And in the days before amplifiers were possible, that milliwatt had to directly drive the other telephone's receiver. So, as much of that milliwatt as possible had to get to the receiver. That meant matched impedances.

Broadcasting, sound recording and sound films imported telephone practice, including the idea of matched impedances. 600 ohms became the universal professional standard for all inputs, outputs and transmission lines. At the time, there was a second good reason for matching everything: vacuum tube amplifiers behave unpredictably unless they operate into the specific load impedance they're designed for. Moreover, radio during its "golden age" involved remote broadcasts on a daily basis. Making everything 600Ω meant anything could be connected to anything, including telephone lines, on a moment's notice.

Studios were wired with incredibly complex networks designed to insure that no piece of equipment would ever see a source or load that wasn't 600Ω. Cabling was brought to **patch bays**, rows of connectors wired to every input and

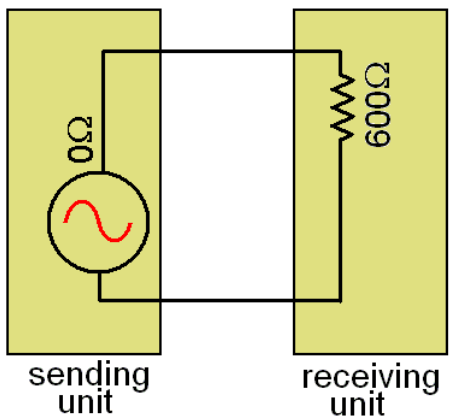


output of every piece of equipment. When an output was not patched to any input, contacts in the connectors, called **normal** contacts, would insert a 600Ω resistor in the circuit.



But this is not always the best way to do things, especially today. Transistor and IC amps don't care what load impedance they drive (within limits), and they're so cheap there's no reason to care about maximum power transfer; any loss can easily be made up. Rather than going for *maximum* power transfer, it is often more important to go for *most efficient* power transfer.

Maximum Efficiency



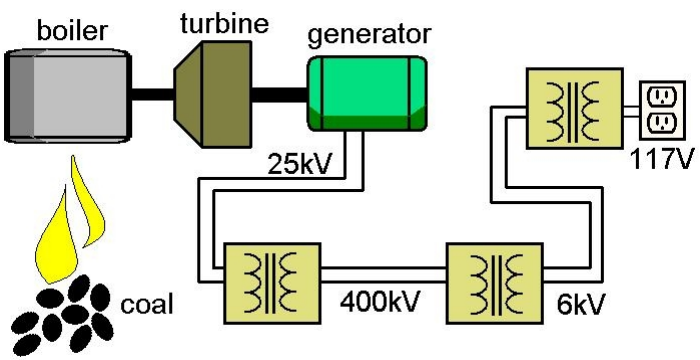
The greatest power transfer efficiency – the least power wasted – occurs when the source impedance is zero. Power, as you know, is voltage times current. When a load is connected to a voltage source, the same current flows in both. If there is no internal resistance in the source, there can be no voltage drop, so there is zero power dissipation in the source. But the load impedance cannot be zero, or it would short out the source. Thus a system with better than 50% efficiency cannot be a matched impedance system.

EM2 Chapter 10 page 19

There's a story told that Thomas Edison's generators were 90% efficient, and he was visited by a group of professors who told him he was doing it wrong; he should make his generators 50% efficient. Edison laughed and told them he was working on a generator that would be 95% efficient.

What would a 50% efficiency have meant? If the power plant put a megawatt out onto the power lines, it would also have burned up another megawatt in the plant. Surely that's not the optimum situation.

Matching is even more inefficient when there are many components between source and load. For example, in a power system, the power plant might burn coal. Heat becomes steam, which turns a turbine, which turns a generator, which puts power onto high-voltage power lines, which is transformed to 117V at your house. If every one of those steps were 50% efficient, just a few percent of the coal's heat would come out as usable power.



So why do people still think in terms of matched impedances? One reason is habit. A second reason is that there are still cases where maximum power transfer is important. For example, the signal from a receiving antenna is often extremely weak, and it's necessary to make best use of it. So, partly for this reason, radio antenna connections are made to match the antenna impedance.

The third and best reason has to do with the way signal moves along a cable. In many cases, especially at high frequencies, the cable's characteristics must be taken into account. If the interface at either end is wrong, the signal gets screwed up.

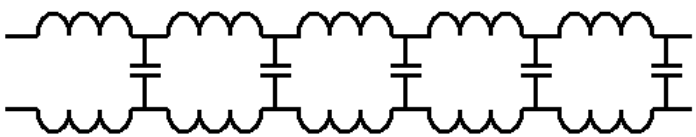
TRANSMISSION LINES

If a cable is long enough for the signal to travel along it in waves, it becomes not just a cable but a **transmission line**. The sending unit is far enough from the receiving unit that it does not “see” the receiving unit's input as its load. As far as it is concerned, the transmission line is the load. Likewise, the receiving unit does not “see” the sending unit as the source of the signal. As far as it is concerned, the transmission line is the source.

A transmission line is a cable that, due to its length, has become such a significant part of the system that its characteristics and requirements join the list of factors that must be considered if the system is to work properly.

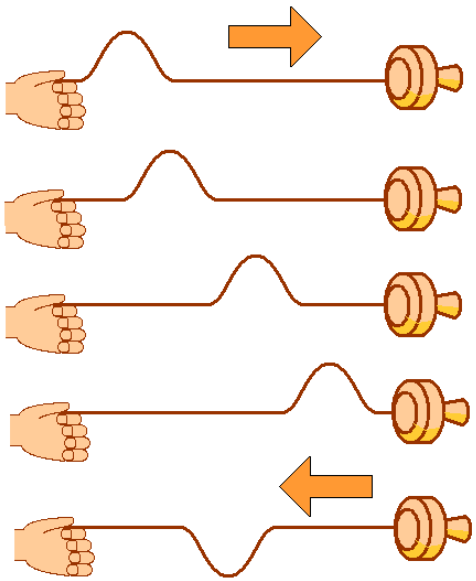
A transmission line has a **characteristic impedance** that has little to do with the resistance of its conductors. You can't measure this impedance with an ohmmeter, because it's an AC phenomenon. The best analogy I can think of is to compare it to sound waves moving through air or water, which have certain densities, certain amounts of mass per cubic foot, which affect how waves travel through them. The different densities of air and water are like cables with two different impedances.

A transmission line can be thought of as a bunch of series inductors and shunt capacitors along the length of the cable.



We know that inductors and capacitors have reactances, which is to say impedances, and this gives the cable an overall impedance.

Whenever waves travel through a medium, if there is an abrupt change in density – in electronics, a change in impedance – some of the wave will be reflected back toward the source. You can watch this happening. If you take a rope, tie it to a doorknob, then stand with the other end in your hand and stretch the rope, you have a line that you can send waves down. Jerk your hand up or down, and you'll see a wave travel down the rope, bounce off the door handle, and travel back to you.



In a transmission line, the same thing happens to the signal whenever it encounters a change in impedance. When the signal gets to the end of the line, if the impedance of the receiving unit doesn't match the line impedance, part of the signal will be reflected back to the source. If the source also doesn't match the line, part of the signal will be reflected again, toward the load.

This means every wave will arrive several times at the load. If the reflections are bad enough, everything will get jumbled up and the signal will become useless. Think of a cathedral where someone is talking, and there are so many sound reflections you can't understand what they're saying.

If you've dealt with whole-house video systems, you may have observed that if any TV is disconnected from the system, anywhere in the house, and not replaced with an equivalent imitation load, the picture is degraded everywhere else. This can be avoided if special precautions are taken, but the fact that you have to take them shows how detrimental reflections can be.

Connecting a matched impedance to a transmission line is called **terminating** the line. For example, video coax is 75Ω, and the inputs and outputs of video equipment are 75Ω to properly terminate the lines.

When is a Cable a Transmission Line?

A cable is a transmission line if it's long enough for reflections to develop. If it's shorter than that, it's just a wire with voltage on it, and it doesn't really have a characteristic impedance. The source “sees” the load directly.

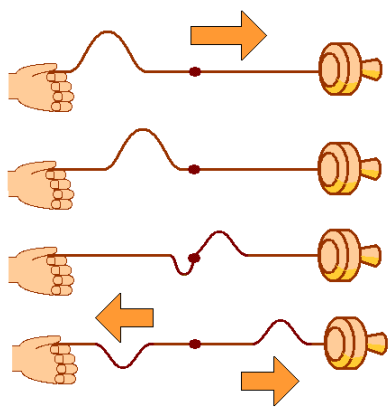
Reflections begin to develop when the cable is about $^1/_{10}$ wavelength long. You remember that wavelength is inversely proportional to frequency; the higher the frequency, the lower (shorter) the wavelength. A 1MHz AM station has a wavelength of 300 meters, or about $^1/_5$ of a mile. For a 100MHz FM station, it's 3 meters, or about 10 feet. At 3GHz it's 4 inches. So, at 1MHz a cable starts to become a transmission line at 100 feet. At 100MHz the dimension is 1 foot, at 4GHz less than half an inch. This means at 4GHz even the printed traces between components on a PC board are transmission lines.

To put it into a formula, a cable is a transmission line if:

$$L > 3 \times 10^7 / f \quad \text{length in meters}$$
$$L > 10^8 / f \quad \text{length in feet}$$

L = length
f = frequency in Hz

When the system gets complicated, keeping things matched gets complicated. You can't distribute a signal to six different places by splicing six different wires together. That puts a “lump” in the impedance partway down the cable.



That's equivalent to tying a knot in the rope of our earlier example. The knot, that is the splice, is an abrupt change in impedance, and part of the signal will be reflected back toward the source.

The proper way to split a signal out to several lines is to insert a **splitter** to insure that every part of the system is properly terminated. If one of the outputs is not used, e.g. a 6-output splitter used to feed 5 lines, the unused output should be terminated with a **dummy load** to avoid screwing up the other lines.

Of course, when you split a signal, the level goes down. This loss is made up by using a **distribution amplifier**, which has enough gain to offset the splitting loss. Some distribution amplifiers have fully isolated outputs, so if one line is not properly terminated, it doesn't affect the other lines.

And that, ladies and gentlemen, is the 21st century reason for matching impedances. A mismatch seldom shows up as a lack of power or a misbehaving piece of gear. It shows up as reflections that corrupt the signal. A technician can waste a lot of time trying to track down a malfunction in a unit that isn't malfunctioning, but has been connected improperly.

Earlier, I said that impedance matching is obsolete in audio. At audio frequencies, a cable has to be a mile long before it's a transmission line. But there's one part of audio where impedances must still be matched. The wired (land line) telephone system is full of lines long enough to be transmission lines. Also, signals are sent both directions over a telephone line, and the impedances have to be right to keep them from getting mixed up.

FYI, telephone lines have 48V of DC on them. If you ever have to connect to a telephone line, be sure you follow all the rules the telephone company gives you.

HIGH FREQUENCY LOSS

All cables attenuate high frequencies due to the series inductance and shunt capacitance described earlier. Unfortunately, shielded cables are worse. The higher the frequency, the greater the loss.

With RF signals, this shows up as simple attenuation. If all the frequency components are close to each other, all you have to worry about is loss of level, and where to insert an amplifier to make up the loss. For example, an FM signal is 200kHz wide. For a station transmitting at 100MHz, that's a 0.2% bandwidth. Whatever the attenuation may be, it will be about equal for every frequency within that channel. That makes our problem a simple one of amplification.

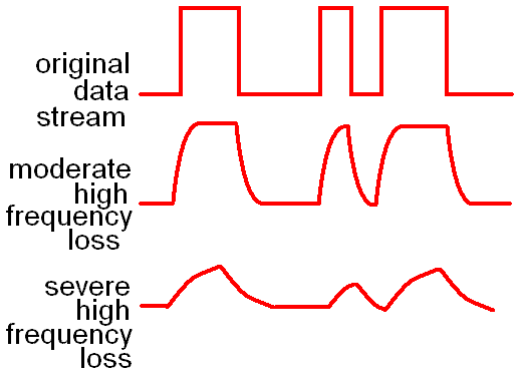
If we're dealing with a receiving antenna picking up a weak signal, it may be necessary to place an amplifier right at the antenna to boost the signal before the cable gets a chance to attenuate it. This is normal practice with dish antennas, which work with such weak, high-frequency signals that cable attenuation is a serious problem.

The situation is worse if the frequency span is great. For example, the signal that comes out of a TV camera occupies a spread from 30Hz to at least 4MHz, much more for hi-def TV. The problem is that cable will not attenuate all of those frequencies equally. Loss will be greater toward the high end of the spectrum, which will result in a picture with blurred edges and washed-out color. This means the signal may need to be processed by correction circuitry with a frequency response that's the opposite of the cable's attenuation characteristic. This is called **equalization** and we'll study it in EM3.

Steps to take:

- Use low-loss cable. In general, fat coax is lower loss than thin coax.
- Try to route the cable so it's as short as possible.
- Equalize if necessary.

Digital data tends to look like square and rectangular waves; high frequency loss shows up as rounding or smoothing of the corners. In this regard, there really is a difference between good name-brand cable and cheap stuff. It's probably bad form to promote a particular manufacturer in a study guide, but in my experience, Belden cable tends to work where other manufacturers' cable won't.



SUMMARY

To get a signal from one unit to another with minimum degradation:

- Be sure every signal is referenced to the proper ground.
- Be sure the signal is protected from electrostatic and electromagnetic interference.
- If the cable is long enough to form a transmission line, be sure the source and load impedances are as close as possible to the cable impedance.
- If frequencies are high enough or the cable is long enough to cause high-frequency loss, minimize the loss by running the cable by the shortest possible route and using low-loss cable.

We've studied signal degradation and ways to avoid it enough for the time being. Now let's look at cable types.

COMMON CABLE TYPES

We're going to look at four kinds of cable: single-conductor shielded, paired shielded, paired unshielded, and twinlead.

SINGLE-CONDUCTOR COAX

“Single conductor coax” is a confusing name; obviously, if you have a signal wire surrounded by a shield, that's two conductors. When people use the expression, they are referring to the number of signal wires, not counting the shield.

General Purpose

For transporting signals over short distances, the choice of cable depends on convenience, availability and cost. Most cables have between 22- and 26-gauge center conductors, and foil or braid shields.

Heavier-gauge signal wires are a waste of money unless the cable will be carrying high-power signals, transporting signals over long distances, or supplying very low-impedance loads. Some flexible cables, like those used for microphones, have 18-gauge conductors, but that's for strength, not electrical properties.

RF and Video Coax

RF/video coax is designed to minimize high frequency losses. Better cables have double or quadruple shielding to improve interference rejection.

As a general rule, the larger diameter the cable, the less loss. 75Ω cables are used for video; other RF applications tend to use 50Ω cable. One reason for that is that a 50Ω cable is a good match for many types of antennas. This eliminates the need to feed the signal through a balun at the antenna, with its inevitable losses.

Here are the most common RF cable types.

TYPE	*Z*	*DIAMETER*	*LOSS*
RG-6	**75Ω**	**medium**	**medium**
RG-8	**50Ω**	**medium**	**medium**
RG-11	**75Ω**	**large**	**low**
RG-58	**50Ω**	**small**	**high**
RG-59	**75Ω**	**small**	**high**

Bundled

Some cables have multiple single-conductor coaxes, along with twisted pairs and single conductors. VGA cables, for example, have separate coaxes for red, green and blue video and for horizontal and vertical synchronization, plus individual wires for other functions, all enclosed in an overall shield.

SHIELDED, PAIRED

The most common cable of this type is a single pair inside a shield. If the cable has multiple pairs, each pair may be individually shielded, all the pairs may be enclosed in an overall shield, or both. If the pairs are individually shielded, the shields may be in contact with each other or insulated. If these pairs are split out to different units in an installation, it's sometimes a challenge to figure out the best places to ground the shields. Before you buy the cable, draw the system on a piece of paper, identify the proper reference for each signal, and choose the ground point or points.

UNSHIELDED, PAIRED

The trend in recent years has been toward unshielded pairs, as the frequencies associated with data and high-definition video have risen. Due to its lower capacitance, unshielded pairs have lower loss at high frequencies. The use of unshielded pairs means signals must be well balanced and rise and fall times must be controlled to avoid excessive radiation. Sometimes a bundle of pairs will have an overall shield, which improves the situation to a degree. Since the overall shield is not wrapped tightly around any particular pair, it does not greatly increase capacitance.

There is a color code that was established many years ago in the telephone industry. This code is widely (but not universally) followed in other applications as well.

- The code accommodates up to 25 pairs.
- Each pair is identified by a major and a minor color. There are five major and five minor colors. They are a bit like decimal numbers in the tens place and the ones place: the code first steps through the five minor colors, then goes to the next major color and again through the five minor colors.
- Polarity is indicated by the color scheme of a particular wire. The “+” wire's insulation is the major color for that pair, with a minor-color stripe on it. The “-” wire's insulation is the minor color for that pair, with a major-color stripe on it. It may seem strange to be speaking of “+” and “-” with AC; but that's the standard way of identifying in- and out-of-phase.

Here are the standard colors. This list is repeated in the "review" section at the end of this study guide, along with selected information from other chapters which you will want to print out and keep.

MAJOR COLORS	***MINOR COLORS***
White	**Blue**
Red	**Orange**
Black	**Green**
Yellow	**Brown**
Violet	**Grey**

Here's how the color code works.

Pair 1, + wire	white with blue stripe
Pair 1, - wire	blue with white stripe
Pair 2, + wire	white with orange stripe
Pair 2, - wire	orange with white stripe
Pair 3, + wire	white with green stripe
Pair 3, + wire	green with white stripe

This goes on through the minor colors until pair 5, white/grey. Then we go to the next major color:

Pair 6, + wire	red with blue stripe
Pair 6, - wire	blue with red stripe
Pair 7, + wire	red with orange stripe
Pair 7, - wire	orange with red stripe

Pair 11 starts with black/blue, pair 16 with yellow/blue, pair 21 with violet/blue. A complete chart with colors for every wire is printed in the review section.

Cables with fewer pairs follow the convention as far as they can. A desk telephone handset has two pairs, white/blue and white/orange. A CAT5 cable has four: white/blue, white/orange, white/green, white/brown.

Larger cables are divided into groups of 25 pairs, identified by loose wraps of thread colored with the minor colors. A 100-pair cable would have 25 pairs wrapped with blue, 25 wrapped with orange, 25 with green, 25 with brown.

Other multi-pair cables follow the resistor color code, or sometimes a code known only to a little man with thick glasses who lives in the basement of the factory and mumbles to himself in a foreign language. Yer on yer own.

For single-pair cables with strange colors, here are the rules I follow:

1. If there's a red wire, that's always +.
2. If there isn't a red wire, + is whichever color is lighter.

TWINLEAD

Twinlead is two wires spaced about ½” apart, separated by a thin sheet of insulation, resembling a builder's I-beam. It's a good low-loss cable where there isn't much risk of interference, which once made it popular for over-the-air reception from TV antennas. But noise sources have multiplied, and long sections of twinlead aren't used much any more. However, short sections are still widely used with FM radios. Its characteristic impedance is 300 ohms, which is not an accident: that's the impedance of a folded dipole antenna, the most common type of simple antenna for FM radio. We'll be talking about folded dipoles in EM6.

PREPARING A CABLE FOR CONNECTION

In EM5, we're going to describe in detail how to prep a cable when we talk about soldering. For now, a few general rules.

1. Don't strip the jacket and insulation back any farther than necessary. That just leaves bare wires exposed, and as the cable is flexed in use, the wires may short.
2. Try not to nick the conductors when you're stripping the insulation. If the cable is flexed, they'll break where you nicked them.
3. When prepping a braided shield, separate the braid and pull it to one side. This will be a lot easier if you didn't strip the jacket back a long way. RF connectors have special cable prep rules. You can find the right ones for each connector on the web.
4. If there is cloth or string in the cable, resist the temptation to burn it away with a cigarette lighter. That will just blacken the conductors and you won't be able to solder to them, and it will create an ugly, nonprofessional mess besides. Just trim them with your smallest side cutters.
5. Hopefully, you will be using new cable with shiny wires. However, if you have to solder tarnished wires, take a knife or fingernail file and scrape them. Tarnished wire can't be soldered.
6. DON'T FORGET TO PUT THE CONNECTOR SHELL OVER THE END OF THE CABLE! Does it sound like I'm speaking from sad experience? And mine isn't the worst: I could tell you about a guy who forgot the connector shell to a TV camera cable, back in the days when a camera cable was maybe eight coaxes and forty other wires. He discovered his oversight just as he was about to connect the last wire.
7. Be sure not to leave any “flying” strands. Murphy being the jerk he is, they are sure to eventually short out the signal. Cut them off close. Better yet, practice until you are skilled enough to not have flying strands in the first place.

I've said nothing about the connectors themselves. They will be covered individually, along with the signals they usually carry, in EM3 through EM6.

Guess what! You're done with EM2. Print out Chapter 11, the Summary, and review it. Put the pages with the charts and equations into the folder you started (right?) with EM1. Take the practice quizzes. Go take the exam and get your ETA EM2 cert.

And hey, if you want to feel smug, go back through EM1 and marvel at how simple it seems to you now.

End of Chapter 10

CHAPTER 11 – EM2 REVIEW

These pages are in a printer-friendly format because we want to encourage you to print them out and go over them until you can describe each of these items in your own words, which is similar to what you'll be asked to do when you take the exams for ETA certification. The various charts and tables from all the chapters are reproduced at the end of this chapter, and you should print those out too so you'll have them handy.

These pages are good for more than just test-taking. You should keep them around and read over them now and then to remind yourself of some things you may have forgotten.

CHAPTER		PAGE
1	Introduction	2
2	Math	2
3	Waves and Waveforms	4
4	Test Equipment and Measurements	7
5	AC Components	9
6	Principles of Capacitance	11
7	Principles of Inductance	12
8	Principles of Transformers	13
9	RC, RL, LC, RLC Circuits	15
10	Cabling	17
---	Charts and Graphs	

CHAPTER 1 – INTRODUCTION

ALTERNATING CURRENT – abbreviated **AC**, current that continuously changes direction, back and forth, usually many times per second.

MIXED AC AND DC – a voltage or current that combines DC and AC. The DC may take the form of a small offset voltage on a large AC voltage, the AC may be a small ripple on a DC voltage, or they may be about equal. With suitable circuitry, they can be separated and dealt with individually.

TIMELINE AND COMPLEXITY – DC power came decades before AC power; DC signals (telegraph) came decades before AC signals (telephone). This is due to the greater complexity of AC signals, and the need for 19th-century technology to advance before it could handle them.

SIGNAL – any electrical transmission whose purpose is to convey information

AUDIO – sound transmitted over wires, always AC, often combined with DC

RADIO – signals transmitted wirelessly through air and space, always AC

KEY – a switch used in Morse code and other on-off systems to make or break a signal circuit

AC VOLTAGE SOURCE – an idealized concept, a source of AC voltage comparable to an idealized battery supplying DC. The idealized source has an internal resistance of zero, meaning that the voltage does not change with changes in load current.

IMPEDANCE – a generalized term similar to resistance, but covering all cases of opposition to current flow, including those where voltage and current are out of step with each other.

CHAPTER 2 - MATH

4-QUADRANT GRAPH – a graph with the origin in the center, and X- and Y-axes extending both directions from it. Quadrants are numbered I, II, III, IV starting with the upper right quadrant and going counter-clockwise.

TRIGONOMETRY – the branch of math dealing with angles and the relationships between the sides of angles.

RADIUS – the distance from the center of a circle to its edge.

DIAMETER – twice the radius, the distance across a circle.

CIRCUMFERENCE – the distance around a circle, if it were broken and straightened out.

Review page 2

PI – the Greek letter π, pronounced "pie," the ratio of a circle's diameter to its circumference. π is an irrational number whose value, truncated to three places, is 3.14. 2π occurs often in electronic calculations; its truncated value is 6.28 or in these tutorials, 6.3.
RIGHT TRIANGLE – a triangle where one of the internal angles is 90 degrees.

SIDES OF A RIGHT TRIANGLE – if we choose one of the internal angles which is not 90 degrees, then from the standpoint of that angle, the short side next to it is the **adjacent**, the short side across from it is the **opposite**, and the long side is the **hypotenuse**. Though the hypotenuse is the longest side, the sum of the lengths of the short sides is always greater.

PYTHAGOREAN THEOREM – if you square the lengths of the short sides, then add them, then take the square root of the sum, you get the length of the hypotenuse. In electronics, the pythagorean theorem can be used with trigonometric ratios to find voltage at any instant. It is also used in calculations in situations where voltage and current are out of step with each other.

TRIGONOMETRIC RATIOS – ratios of the lengths of the various sides, stated as decimal numbers. The most common ratios, and the ones we use, are the sine, cosine and tangent.
Sine = Opposite divided by Hypotenuse
Cosine = Adjacent divided by Hypotenuse
Tangent = Opposite divided by Adjacent
These quantities are often abbreviated **sin, cos, tan**.

PHASOR – once called a **vector**, a rotating quantity represented by an arrow whose "back" end is usually anchored at the origin of an X-Y graph, and whose tip rotates counterclockwise. In electronics, this most often represents a sine wave AC voltage. A phasor can also occupy other positions; often, a phasor will begin at the tip of another phasor to indicate a voltage drop or other quantity.

RADIAN – a section of a circle's circumference that, if straightened out, would be one radius long. 2π radians equal a circle's circumference.

REAL AND IMAGINARY NUMBERS – real numbers are points on the X-axis. Imaginary numbers are points on the Y-axis. In most mathematics, imaginary numbers are preceded by a lower-case "i," but we use that for AC current, so in electronics an imaginary number is indicated by a lower-case "j." Like real numbers, imaginary numbers can be positive or negative.

COMPLEX NUMBERS – combinations of real and imaginary numbers, indicated thus: (6, j3). Complex numbers can be used to specify any point in any quadrant of an X-Y graph.

LINEAR SCALE – a scale where the equal distances between markings represent equal increments, in an absolute sense: 1, 2, 3, 4, 5 and so on.

PROPORTIONAL SCALE – a scale where the equal distances between markings represent equal increments, in a proportional sense: 1, 2, 4, 8, 16 and so on.

Review page 3

LOGARITHMIC SCALE – a scale where the distance between adjacent decimal numbers represents their proportional difference; e.g. the distance between 1 and 2 = the distance between 10 and 20. A logarithmic scale has no zero or negative numbers.

LOGARITHMIC CYCLE – a completely different meaning from the one attached to electrical alternations; a cycle is a 10:1 ratio on a logarithmic scale. A 3-cycle log scale beginning at 10 would have an unbroken span covering the ranges from 10-100, 100-1,000 and 1,000-10,000.

LOG-LINEAR GRAPH – a graph where the X-axis is logarithmic and the Y-axis is linear, or vice versa.

LOG-LOG GRAPH – a graph where both axes are logarithmic.

CHAPTER 3 - WAVEFORMS

AMPLITUDE – also called **level**, the strength of a signal measured in volts, amps or watts. With AC, amplitude is specified as an **absolute number**, taking into account both positive and negative alternations.

PERIODIC – an AC pattern that repeats at regular intervals. Each repetition is a **cycle**. The number of cycles per second is the **frequency**, measured in **Hertz**, abbreviated **Hz**. When a waveform repeats continuously, this is called **oscillation**.

PERIOD – the time occupied by one cycle. Period is the reciprocal of frequency; e.g. a 5Hz frequency corresponds to a $^{1}/_{5}$ second period.

SPECTRUM – the entire range of frequencies from earthquakes to X-rays and beyond. Alternately, refers to the harmonic content of a waveform.

BAND – a subsection of the spectrum.

AUDIO BAND – frequencies in the general range of human hearing, further subdivided. **Sonic** frequencies are those we can hear, roughly 16Hz – 20kHz. **Subsonic** vibrations are lower than we can hear, below 16Hz. **Ultrasonic** vibrations are higher than we can hear, above 20kHz.

RADIO BAND – usually taken to mean frequencies of about 50kHz and higher. Subdivided into a number of smaller bands; see chart at the end of this chapter.

WAVE – another word for cycle.

WAVEFORM – the shape of a cycle when shown on a graph.

RADIATION – the transmission of electromagnetic signals through air and space.

Review page 4

PROPAGATION – the transmission of electrical signals through wires or other media.

SPEED OF RADIATION – the speed of light, 186,000 miles per second. For convenience in calculation, engineers and techs think in metric terms: light and radiated signals travel at almost exactly 3×10^8 meters/second.

SPEED OF PROPAGATION – speed of a signal through a medium, a bit slower than speed of radiation. Actual speed depends on the medium. If the medium is a cable, the manufacturer's website should give the speed of propagation.

WAVELENGTH – the distance in air and space occupied by one wave as it travels along; that is, the distance between a point in one wave and the corresponding point in the next wave.

DUTY CYCLE – the percentage of a wave occupied by the positive alternation. A symmetrical wave would have a 50% duty cycle.

SQUARE WAVE – a wave with flat tops and bottoms, square corners, and a 50% duty cycle. A square wave can be thought of as a DC voltage that changes polarity. In a perfect square wave, the time it takes voltage to rise and fall would be zero. In a real-world square wave, the rise and fall take some time. These intervals are called **rise time** and **fall time**.

RECTANGLE WAVE – like a square wave, but with unequal positive/negative duty cycles.

TRAPEZOID WAVE – a square wave whose rise and fall times have been stretched to the point where they take up a considerable part of the cycle.

TRIANGLE WAVE – a wave whose voltage rises linearly for half the cycle and falls linearly for the other half. A linear rise or fall is called a **ramp**.

SAWTOOTH WAVE – like a triangle wave but with very unequal rise and fall times, e.g. 90% / 10%.

SINE WAVE – a smoothly varying waveform that is the time equivalent, and graphical representation of, circular motion.

PEAK, VALLEY – the top and bottom points of a sine wave

ZERO CROSSING – the moment when a waveform crosses the X-axis, going from the negative to the positive alternation or the reverse.

PHASE ANGLE – when referring to a single sine wave, a point along the wave. A sine wave cycle is divided into 360 degrees, since it corresponds to circular motion, with zero degrees at the positive-going zero crossing. From there, every point along the wave corresponds to a certain number of degrees. For example, the negative-going zero crossing is 180 degrees.

Review page 5

HARMONICS – the series of mathematically related frequencies that are components of a waveform. The **first harmonic**, which is the basic frequency of the wave, is the **fundamental**. Higher harmonics, which are multiples of the fundamental frequency, determine the specific shape of the wave. Harmonics are numbered, and the numbers correspond to multiples of the fundamental frequency, e.g. the 3rd harmonic is 3 x the fundamental frequency. The sine wave is the only waveform that contains only fundamental.

COMPLEX WAVE – a waveform made up of sine waves at harmonically related frequencies. Every waveform except the sine wave is a complex wave.

LOW AND HIGH ORDER HARMONICS – low and high harmonic multiples, respectively. High-order harmonics correspond to fast rise and fall times, which are responsible for the interference caused by such things as lamp dimmers.

SPECTRUM ANALYSIS – the graphical representation of the harmonic content of a wave, or more generally, of signal amplitude vs. frequency across a selected band.

RMS VOLTAGE – the effective voltage of a waveform, obtained by taking many samples along the wave, squaring each sample, averaging the squares, and finding the square root of the average. In practice, this is done in real time by electronic circuitry.

PEAK VOLTAGE – voltage at the peak or valley of a wave. For a sine wave, peak voltage is about 1.4 times RMS voltage; peak power is 2 times "RMS" power, the average power calculated from the RMS voltage.

PEAK TO PEAK VOLTAGE – the entire voltage span of a waveform from its most positive to its most negative point.

TRANSIENT – a voltage spike of short duration, containing energy at every frequency across the spectrum. The usual way of measuring a transient is to measure the height of the peak.

NOISE – transients occurring so frequently that they become a continuous signal. Noise contains energy at every frequency. In the short term, noise is unpredictable, but in the long term, an average RMS voltage can be measured.

NOISE FLOOR – the small amount of noise present in even the quietest electronic system. This establishes the lowest signal level that the system can handle without serious degradation. In many cases, the signal must be kept relatively "clean," which means it can never be allowed to approach the noise floor.

COMPLEX WAVEFORMS – real-world signal waveforms that may seem chaotic, but which can be analyzed into the simpler waveforms presented in this chapter.

Review page 6

AC TEST EQUIPMENT

OSCILLOSCOPE – electronically draws an X-Y graph in real time. This can be done by analog circuitry, or the same functions can be performed by digital circuitry that mimics the analog functions. An electron beam or its digital equivalent travels across the screen left to right. It is deflected up and down to trace the signal waveform. Most of the following test equipment entries refer to oscilloscopes.

STORAGE OSCILLOSCOPE – captures a snapshot of a real-time waveform for later study. The only worthwhile storage scopes these days are digital.

TRIGGER – a circuit that looks for a certain point in the signal waveform to start the beam sweep or digitally anchor the waveform display. This insures that each successive wave will show in the same place on the screen, so the display will be stable.

GRADUATIONS – lines on the display which can be used to manually determine waveform period and amplitude.

RANGE SWITCHES – control vertical sensitivity and horizontal sweep speed. Used to dimension waveforms on the display so they can be studied.

FINE ADJUST – pots which give fine control of horizontal speed and vertical size. Usually, when the pots are set fully clockwise, speed and size are accurately calibrated to the size of the screen graduations. The user should remember to return find adjust pots to these settings.

5X – on analog scopes, a switch that causes the trace to overscan by 5:1. This permits closer observation of a part of the displayed waveform.

HORIZONTAL AND VERTICAL POSITION – used to center the trace on the screen, or in some cases to move it off center.

DUAL TRACE – provided by circuitry which allows the scope to display two signal traces at the same time. In an analog scope with a single beam, this involves switching the beam back and forth between traces. In a digital scope this is not necessary since the display is essentially a video screen showing a picture of whatever the computation circuitry sends it.

DC / AC / GROUND SWITCH – a switch at the signal input which can insert a blocking capacitor in the circuit, so the scope will display the AC component of a signal. The ground position permits easy trace centering.

FOCUS, ASTIGMATISM, TRACE ROTATION – analog scope controls to obtain a sharp, properly aligned trace. Unnecessary in a digital scope.

CALIBRATION – the process of insuring that the period and size of the displayed waveform is accurately matched to the graduations on the scope face.

Review page 7

INPUT LOADING – the load which the signal input places on the circuit being measured. This is the combination of the load presented by the scope and the load presented by the test probe.

10X PROBE – a probe with a 10:1 attenuator built into the tip. This greatly reduces input loading. A 10X probe usually has a switch to remove the attenuator when measuring low-level signals.

RINGING – one of several problems that can be caused by improper probe adjustment or improper grounding. Shows up on the screen as a tendency toward oscillation on transients or waveforms with fast rise/fall times.

ACTIVE PROBE – a probe with a built-in amplifier, used to reduce the load on the unit under test to a minimum. Care must be used to avoid damaging an active probe with a signal voltage that is higher than it can handle.

MEASUREMENT ARTIFACTS – waveform distortions that are actually caused by scope limitations. Rounding of corners or showing square waves as sine waves are caused by the high-frequency limitations of the scope. If the input switch is set to AC, the tops and bottoms of low-frequency square waves will be tilted. The user must know the scope well enough to know which effects are actual waveform distortions and which are artifacts.

Other test gear:

SPECTRUM ANALYZER – displays a real-time X-Y graph where the X-axis corresponds to frequency. The resulting display shows the relative strength of every signal at every frequency within the displayed frequency range.

DISTORTION ANALYSIS – shows the percentage of one or more types of waveform distortion.

INDUCTANCE-CAPACITANCE TESTER – measures inductors and capacitors, and tests for partially shorted or open components.

CURRENT-TRANSFORMER METER – a meter that can be clamped around a wire to inductively measure current without actually breaking the circuit and inserting an ammeter. Most commonly known by the brand name “Amprobe.”

FREQUENCY COUNTER – counts and displays a signal waveform's frequency. Digital oscilloscopes often have frequency counter, spectrum analysis and multimeter functions built in.

SIGNAL GENERATORS – generate continuous oscillations for testing purposes. The most common are resonant sine wave generators, which have the advantage of spectral purity, and function generators, which can generate several different waveforms and have the advantage of complete voltage stability.

Review page 8

NOISE GENERATORS – generate random or pseudo-random noise for test purposes.

WHITE NOISE – noise which contains the same amount of energy in every Hz of bandwidth; e.g. the same amount of energy between 100Hz and 200Hz as between 1,100Hz and 1,200Hz. Useful when displaying spectrum on a linear scale.

PINK NOISE – noise which contains the same amount of energy in every proportional increment of bandwidth, e.g. the same amount of energy between 100Hz and 200Hz as between 1,000Hz and 2,000Hz. Useful when displaying spectrum on a logarithmic scale.

CHAPTER 5 – COMPONENTS

CAPACITOR – stores energy as an electric charge as a potential across two conductive plates.

CAPACITANCE – the amount of charge a capacitor will store per volt of potential, measured in **Farads, abbreviated F**. This is such a large unit of storage that it's impractical for most purposes. Most capacitors are measured in **microfarads** and **picofarads**, abbreviated **µf** and **pf**.

PLATES – the conducting surfaces between which the charge is stored. Though they are called “plates,” they often take the form of long strips of foil rolled up into cylinders.

DIELECTRIC – the nonconducting material between the plates. In **dry** capacitors, it may be ceramic, mica, a plastic film, air, occasionally glass. In **electrolytic** capacitors it is a thin insulating layer that is chemically deposited on one of the plates.

NONPOLAR – dry capacitors do not have a polarity; they can be connected into a circuit either way.

POLAR – electrolytic capacitors have a polarity; if connected into a circuit so there is a DC voltage across them in the wrong direction, they will be destroyed. A nonpolar electrolytic can be made by connecting two electrolytics back to back.

BREAKDOWN – when the voltage across a capacitor exceeds the dielectric's ability to withstand it, a spark jumps between the plates. In an air-dielectric capacitor, this may not cause damage, but in a solid-dielectric capacitor, the dielectric is destroyed and the capacitor must be replaced.

SUPERCAPACITOR – a comparatively new kind of capacitor which stores much more charge; even small supercapacitors may have several Farads worth of storage. At the moment, supercapacitors are limited to a few volts maximum.

AXIAL LEADS – leads located on the axis of a tubular capacitor, one from each end.

Review page 9

RADIAL LEADS – leads located off the axis, both coming out one end of the capacitor.

VARIABLE CAPACITOR – a capacitor whose value can be changed. Widely used as tuning devices in radios.

TRIMMER / PADDER – a small variable capacitor intended for infrequent adjustment

GIMMICK – a low-value capacitor made of a pair of twisted wires.

CAPACITIVE REACTANCE – a capacitor's opposition to current flow. Its symbol is $\mathbf{X_C}$ and it is measured in **Ohms**. It differs from resistance in that a capacitor does not dissipate power, but stores and releases it. Current through a capacitor and voltage across it are not in step with each other. X_C varies inversely with frequency.

INDUCTOR - stores energy in the magnetic field around a coil. The symbol for inductance is **L**. Inductors are nonpolar; they can be connected either way.

INDUCTANCE – the amount of energy an inductor will store per amp of current, measured in **Henrys**, abbreviated **H**.

CORE – what's in the middle of the coil. The three most common materials are air, or a nonconducting, nonmagnetic material like plastic which is the same as air for magnetic purposes; soft iron, or a ferrite material that consists of iron or iron oxide particles suspended in a ceramic material.

EDDY CURRENT – a circulating current induced in an iron core which wastes power.

LAMINATIONS – thin soft iron sheets coated with an insulating material, which are stacked to form the core of an iron-core inductor. The purpose is to reduce eddy currents.

VARIABLE INDUCTOR – an inductor whose inductance can be changed. The most common form involves a core that can be partially removed from inside the coil.

SATURABLE REACTOR – a reactor in series with a load which can be deliberately saturated to reduce its reactance.

MICROPHONICS – momentary changes in inductance caused by vibration transferred to an inductor from an external source, which distorts the shape of the inductor.

SELF INDUCTANCE – the inductance created by an inductor acting independently.

MUTUAL INDUCTANCE – the inductance created by inductors sharing a magnetic field.

INDUCTIVE REACTANCE – an inductor's opposition to current flow. Its symbol is $\mathbf{X_L}$ and it is measured in Ohms. X_L does not dissipate power but stores and releases it.. Current and voltage are not in step with each other. XL varies directly with frequency.

Review page 10

TRANSFORMER – transfers electrical power from input to output magnetically, with no electrical connection. In the process, it may change the voltage-to-current ratio, so output power is eqaul to input power (minus insertion loss) but in a different form.

CORE – the magnetic material on which the transformer is built. Low-frequency transformers use soft iron cores; high frequency transformers use ferrite cores. Very high frequency transformers may be air core.

WINDING – an input or output coil wound o the core.

PRIMARY – the input coil.

SECONDARY – the output coil.

TURN – a loop of wire in a winding. A winding may have just a few turns or thousands.

TURNS RATIO – the ratio of the number of turns in the primary winding to the number of turns in the secondary.

STEP UP / DOWN RATIO – the ratio of input to output voltage. Neglecting losses, this is the same as the turns ratio.

REPEAT COIL – another name for an audio transformer with a 1:1 turns ratio.

ISOLATION – refers to the fact that the primary and secondary are not electrically connected.

CHAPTER 6 – PRINCIPLES OF CAPACITANCE

ENERGY STORAGE – energy is stored as a charge difference across conductive plates. Since this is a stationary charge, it can be stored for a significant period of time.

CHARGING / DISCHARGING – voltage across the capacitor increases as electrons flow into one plate and out of the other. Reversing the flow causes voltage to decrease.

CHARGE / DISCHARGE RATE – the voltage change per second is equal to charging/discharging current divided by the capacitance in farads: **dV/dT = I/C**

TIME CONSTANT – in an RC circuit, the time it takes for a capacitor to reach 63% of charge or discharge: **T = RC**, time in seconds, R in ohms, C in farads.

FULL CHARGE – an arbitrary estimate. A capacitor reaches “pretty much” full charge (or discharge) in five time constants.

COUPLING CAPACITOR – a capacitor inserted between two circuits which passes an AC signal from one to the other but blocks any DC voltage that may be present.

Review page 11

BLOCKING CAPACITOR – another name for coupling capacitor.

DECOUPLING CAPACITOR – usually, a capacitor connected between a power supply rail and ground, which shunts (shorts out) any AC voltage that may be present on the rail. The purpose is to keep signal voltages from leaking from one part of the circuit to another via the power supply rail. Another name for a decoupling capacitor is a **bypass capacitor**.

CAPACITIVE REACTANCE – a capacitor's opposition to AC current flow, measured in Ohms. The symbol is $\mathbf{X_C}$. Capacitive reactance varies inversely with frequency; at any particular frequency, it can be found with the formula: $\mathbf{X_C = 10^6 / 2\pi fC}$, where f = frequency in Hz, C = capacitance in μf.

VOLTAGE / CURRENT RELATIONSHIP – when an AC voltage is connected across a capacitor, the waveform of the current through the capacitor will lead the voltage waveform by 90 degrees.

SCALING – a math-lite way of calculating X_C. At 1kHz, X_C of an 0.016mfd capacitor is 10kΩ. By shifting values up and down, X_C at any frequency for any capacitor can be found. Note that, as either frequency or capacitance goes up, X_C goes down.

SERIES CAPACITIVE REACTANCES – capacitive reactances in series add, like resistances, e.g. the X_C total of three capacitances in series is $X_{C1} + X_{C2} + X_{C3}$. Note, however, that connecting several capacitors in series *reduces* total **capacitance**. This is because capacitive reactance is inversely proportional to capacitance.

PARALLEL CAPACITIVE REACTANCES – the total capacitance of parallel-connected capacitors is the sum of the individual capacitances. Total capacitive reactance, however, is lower than any individual X_C and is found using the product-over-sum or sum of reciprocals formula. Again, this is because capacitive reactance is inversely proportional to capacitance.

PARASITIC INDUCTANCE AND RESISTANCE – real world capacitors aren't pure capacitance. They have a small amount of resistance and inductance, and this changes their behavior at high frequencies, especially for larger capacitors. It would be a good idea to go to the Digi-Key or Mouser website and find specification sheets for capacitors, so you can study these effects.

CHAPTER 7 – PRINCIPLES OF INDUCTANCE

INDUCTOR – stores energy in the magnetic field generated by current flowing through a coil. Since this storage depends on continuously flowing current, an inductor is usually thought of as a momentary energy storage device.

Review page 12

VOLTAGE / CURRENT RELATIONSHIP – placing a voltage across an inductor causes current to change. A constant voltage will cause current to increase/decrease at a constant rate. The rate is proportional to voltage and inversely proportional to inductance: **dI/dT = V/L**. When an AC voltage is applied across an inductor, current lags behind voltage by 90 degrees.

BACK EMF / BACK VOLTAGE – the self-generated voltage that opposes applied voltage across an inductor. Back EMF is the reason that current through an inductor does not change in step with voltage, as it does with a resistor.

FLYBACK – a voltage spike that an inductor generates when it is abruptly disconnected from its circuit while current is flowing.

TIME CONSTANT – in a circuit with a resistor, the time it takes for current to reach 63% of its final level. **T = RL**, where T= time in seconds, R = resistance in Ohms, L = inductance in Henrys.

FULL CURRENT = an arbitrary estimate. In an RL circuit, current reaches "pretty much" its final value in five time constants.

INDUCTIVE REACTANCE – an inductor's opposition to AC current flow, measured in Ohms. The symbol is XL. Inductive reactance varies directly with frequency; at any particular frequency, it can be found with the formula: $\mathbf{X_L = 2\pi fL}$, where L is in Henrys.

SERIES INDUCTANCE – an inductor in series with a load will pass DC but attenuate AC.

SATURATION – complete magnetization of an iron core. When a core is saturated, inductive behavior essentially disappears. At that point, impedance is not much more than the DC resistance of the coil.

MUTUAL INDUCTANCE – occurs when two inductors are physically close enough that their magnetic fields partly overlap. If an inductor's field is generated by an AC current, the field strength will vary as well, and this will induce a voltage in the second inductor. In this way, power can be magnetically transferred between inductors. This is called **transformer action**.

PREVENTING MUTUAL INDUCTANCE – mutual inductance between can be canceled by orienting inductors at right angles to each other, or turning them 45 degrees in the same direction.

CHAPTER 8 – TRANSFORMERS

TRANSFORMER – a device which uses mutual inductance to transfer electrical power from its input to its output with no electrical connection. It can be constructed so as to change the ratio of voltage to current in the process. Output power equals input power, minus insertion loss (less than 100% efficiency).

Review page 13

REFLECTED LOAD – when a load is connected to the secondary of the transformer, it is "reflected" back through the transformer and loads the voltage source. Ideally, when a transformer is not connected to a load, the load that the transformer places on the source by itself will be insignificant.

FREQUENCY LIMITS – a transformer does not work with DC. At lower AC frequencies, a limit is reached where it does not work well. This is due to saturation of the core. At high frequencies, stray inductance and core losses reduce efficiency. Thus a transformer only works efficiently over a limited band of frequencies. The wider the frequency band required, the more difficult and expensive to design and build the transformer; thus transformers are usually made to operate over just the frequency band required.

TAP – a connection to a transformer winding at some point other than the end. Voltage at any point in the winding, as a percentage of the total voltage across the winding, is proportional to the number of turns, as a percentage of the total number of turns.

CENTER TAP – a tap connected at the halfway point in a winding. Equal voltages will appear between each end of the winding and the tap.

AUTOTRANSFORMER – a transformer with just one winding. A tap is connected at some point along the winding to obtain voltage step-up or step-down. For a given power level, an autotransformer is smaller than a regular transformer, but it does not offer isolation.

SPLIT AND MULTIPLE WINDINGS – a transformer can have several primary and/or secondary windings. This may be done to provide several voltages from one transformer, or to allow a choice of voltages. A typical example would be a transformer with a split primary winding that can be connected to work with either 120 or 240 volts.

TOROID – a transformer with a donut-shaped core. Toroids are smaller, lighter and more efficient than equivalent rectangular transformers, and radiate less stray magnetic field. They are much harder to construct and so are more expensive.

IMPEDANCE RATIO – in a signal transformer, the ratio of input to output impedance is the square of the voltage ratio. A signal transformer's performance specifications only apply when the transformer is connected to specified source and load impedances. If it is a 1:1 transformer, source and load impedances will be equal. If it is a step-up or step-down transformer, the impedance ratios will be adjusted accordingly. For example, a 2:1 stepdown transformer might have specified source and load impedances of 600W and 150W respectively.

SHIELDING – a transformer can pick up stray magnetic fields. To prevent this, the transformer may be encased in a soft-iron enclosure. This is called a magnetic shield, and is usually made of **mumetal** for low-frequency transformers.

POWER TRANSFORMER – a transformer built to handle 60Hz AC power, or in some countries, 50Hz.

Review page 14

AUDIO TRANSFORMER – a transformer built to handle audio frequencies.

MODEM TRANSFORMER – an audio transformer with poor low frequency response, acceptable because the data involve does not contain low frequencies.

SWITCHING POWER TRANSFORMER – a power transformer built to operate, typically, at 50kHz or higher.

RF TRANSFORMER – a transformer built to handle high frequencies, in some cases up to GHz. These include antenna, and cable transformers.

CURRENT TRANSFORMER – part of a unit that measures AC current in a wire.

CHAPTER 9 – RC, RL, LC, RLC CIRCUITS

PHASOR DIAGRAM – a convenient way of showing voltage, current and phase relationships in RC, RL, LC, RLC circuits.

SERIES RC CIRCUIT -

- voltages across R and C add up to more than the applied voltage
- voltage across C is 90^{O} out of phase with voltage across R
- voltages across R and C are found using Pythagorean theorem
- in a series circuit, multiple resistances and multiple capacitances can be lumped into a single resistance and a single capacitance
- total impedance is found using Pythagorean theorem (see chart and equation pages)
- current is the same everywhere
- current through C leads voltage across it by 90^{O}
- power may be apparent, corresponding to total power in the circuit, or real, corresponding to dissipation in the resistor
- Z = V / I

SERIES RL CIRCUIT

- works exactly like a series RC circuit, except that current through L lags voltage across it by 90^{O}

PARALLEL RC OR RL CIRCUIT

- Voltage is the same across both components
- Current in R and C or L are found indepenently using Ohm's Law
- Total current in is found using Pythagorean theorem
- Impedance is found by dividing voltage by current: Z = V / I

Review page 15

SERIES LC CIRCUIT

- reactances of C and L are opposites; subtract the smaller from the larger to find net reactance
- the frequency where $X_L = X_C$ is called "f." At this frequency, net reactance = 0. Z = whatever parasitic resistance is part of the circuit.
- at frequencies far below f, net X approximately equals X_C. At frequencies far above f, net X approximately equals X_L.
- at f, the voltages across C and L can be hundreds of times higher than the applied voltage.
- at frequencies other than f, current is found in the usual manner by dividing V by the net reactance. At f, current is voltage divided by parasitic resistance.

PARALLEL LC CIRCUIT

- currents through L and C are opposites; subtract the smaller from the larger to find net current
- at f, current = 0 and Z = nearly infinity
- at frequencies far below f, net X approximately equals XL. At frequencies far above f, net X approximately equals XC.

RESONANCE – the "ringing" behavior of an LC circuit caused by energy being traded back and forth between L and C. An LC circuit's resonant frequency is f, where $X_L = X_C$. Formulas reproduced in last pages of this review.

CHANGING RESONANT FREQUENCY – change either L or C. The change in frequency is the square root of the amount of change in L or C.

RLC CIRUIT – an LC circuit with added resistance, which makes resonant behavior less pronounced.

"**Q**" - quality of resonant circuit. A high-Q circuit is sharply "tuned" to f; a low-Q circuit is broadly tuned and less resonant. Q is the ratio of either X_C or X_L at the resonant frequency to R: **Q = X / R**

DAMPING – the resonance-reducing effect of R. Damping is the reciprocal of Q.

FILTER – a circuit built to discriminate among frequencies, passing some and attenuating others.

PASSBAND OR PASS BAND – the range of frequencies which the filter passes.

ATTENUATION BAND – the range of frequencies which the filter attenuates.

STOP BAND – the attenuation band in a filter where attenuation is very large, such that frequencies in this range are almost totally blocked.

Review page 16

TURNOVER FREQUENCY – also called **transition frequency** or **cutoff frequency**, the frequency which divides the pass band from the attenuation band. This is also called the **3dB down point** because attenuation at this frequency is 3dB.

HIGH PASS – a filter that passes high frequencies and attenuates low frequencies. The attenuation band is DC to the turnover frequency. The passband is from the turnover frequency up.

LOW PASS – a filter that passes low frequencies and attenuates high frequencies. The passband is DC to the turnover frequency. The attenuation band is from the turnover frequency up.

BANDPASS OR BAND PASS – a filter that passes a band of frequencies and attenuates frequencies outside this band. A wide bandpass would have separate low and high turnover frequencies. A narrow bandpass would have a single center frequency, and would pass a defined band of frequencies on either side of the center frequency.

BAND STOP – a filter that passes all frequencies except those in a define band. Typically used to remove a single objectionable signal or group of signals.

ALL PASS – a filter that passes all frequencies, but changes the phase of some relative to others.

L, T, π FILTERS – circuit arrangements of inductors and capacitors that resemble those letters. T and π filters are used where sharper division between pass and stop bands is desired than can be obtained with L filters.

MATCHED IMPEDANCE SYSTEM – a system where the signal source's impedance, the load impedance, and ideally the filter impedance are all the same.

NON-MATCHED IMPEDANCE SYSTEM – a system where the source supplying signal to the filter has zero impedance and the load has infinite impedance. RC and RL filters are most easily designed to work in a non-matched system, but many LC and RLC filters require matched impedances.

CHAPTER 10 – CABLING

CABLE – two or more wires held together and intended to be treated as a unit

JACKET – the insulating outer coating or container that encloses the wires in a cable

ZIP CORD – also known as lamp cord, two wires in two jackets joined in the center by a thin web. The two wires can easily be pulled apart to facilitate connection.

Review page 17

SIGNAL DEGRADATION – as a signal travels through a cable, some modification is inevitable. Assuming the signal was correct to begin with, modification should be considered damage. The science of cabling is largely an effort to minimize this damage.

NOISE – in this context, electrical interference of any kind; anything added to the signal that does not belong there. This noise may consist of the kind of noise defined in chapter 3 – random energy from molecular vibration – or any other kind of interference, such as lamp dimmer buzz. Noise pickup can be minimized by:

- sending a "hot," high-level signal
- using shielded cable
- using balanced wiring
- crossing other cables at right angles
- being careful with grounds
- keeping your distance from interference sources

SIGNAL-TO-NOISE RATIO – also known as S/N, the margin between the level of a signal and the level of the noise that is part of it. Usually measured in decibels.

ELECTROSTATIC COUPLING – coupling noise from a noise source into a signal wire through electrostatic means; a capacitance-based energy transfer.

ELECTROMAGNETIC COUPLING – coupling noise from a noise source into a signal wire through electromagnetic means; an inductance-based energy transfer.

SHIELDED CABLE – a cable where the signal wire/s are wrapped in a conductive foil or braid, which is connected to equipment ground. This provides a zero-potential environment for the signal wires. Shielding is effective against electrostatic noise but not electromagnetic noise.

BALANCED WIRING – the process of sending the signal as an AC voltage across a pair of wires. The equipment on either end is designed to produce, and sense, only voltage differences. Noise, which tends to be the same on both wires, is canceled. A voltage across the wires is called a **differential-mode** voltage. A voltage that is the same on both wires is called a **common-mode** voltage.

UNBALANCED – a signal that is not in the form of a voltage difference between two wires. Most equipment has unbalanced circuitry inside, meaning the signals appear as voltages between points in the circuit and the internal ground. Another term for unbalanced is **single-ended**. Where interference pickup is not a problem, there is no benefit to balanced wiring and unbalanced is cheaper.

TWISTED PAIR – a pair of wires that are twisted around each other, making it almost impossible for an external noise source to affect one differently from the other and create a differential voltage across them. The pair may be shielded or unshielded.

BALUN – a transformer that converts a signal from balanced to unbalanced or vice versa.

Review page 18

FLOATING – a balanced pair that is isolated by transformers from circuitry at either end. This configuration gives maximum common-mode rejection.

RIGHT-ANGLE CROSSING – if a signal cable must cross another cable which may be an interference source, it should cross at right angles.

GROUND POTENTIAL – a voltage that exists between two ground points. If this voltage becomes part of the wiring system, it can become noise added to the signal.

GROUND LOOP – a loop formed when the grounds of two pieces of equipment are connected together by two or more separate ground wires.

COMMON GROUND – using the same ground point as a reference throughout a system, which minimizes certain kinds of ground-related problems.

EARTH GROUND – a connection to the Earth that is often required for safety reasons, but is of little benefit and may be harmful from a signal standpoint.

DISTANCE – the strength of electric and magnetic fields falls rapidly with increasing distance. Interference problems will be greatly reduced if signal lines are kept away from, for example, AC wiring.

TRANSMISSION LINE – what a cable becomes when it is long enough for signals to travel along it in waves. This begins to happen at $1/10^{th}$ of the wavelength of the signal it is carrying.

HIGH FREQUENCY LOSS – cable attenuates high frequencies. The higher the frequency and the longer the cable, the greater the attenuation. This can be minimized by routing cables via the shortest paths possible and using low-loss cable. Unshielded twisted pairs have the advantage of lower impedance; thus they are better for very high frequency transmission if the interference problem can be managed.

CABLE IMPEDANCE – a measure of the characteristic impedance of a transmission line. RF cables tend to be 50Ω, video cables tend to be 75Ω. A flat, unshielded video and FM radio cable called **twinlead** is 300Ω.

IMPEDANCE MATCHING – connecting a transmission line to a signal source with an output impedance and a load with an input impedance that are the same as the transmission line's impedance. Another word for connecting a matched load impedance is **termination** of the line.

REFLECTION – when a transmission line is not properly terminated, some of the signal is reflected back toward the source. Severe reflections can degrade the signal to the point where it is useless.

MAXIMUM POWER TRANSFER – occurs when source and load impedances are matched. This is only a consideration where signal power is very small and best use must be made of it.

Review page 19

MAXIMUM EFFICIENCY – occurs when the source impedance is much lower than load impedance. In general, best efficiency happens when source impedance is as low as possible.

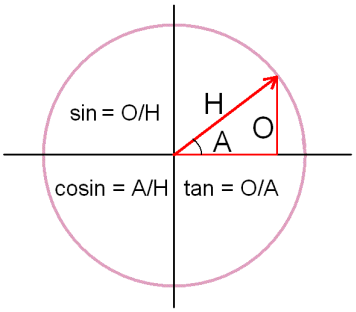
CABLE PREPARATION –

- strip no more insulation than necessary
- cut off any strings or fibers in the cable, don't burn
- if the shield is braided, separate the braid, pull sideways, twist to form a neat bundle
- remember to put the connector shell on before you solder the wires to the connector
- do not leave “flying” strands of wire
- take care not to nick the conductors, they'll break.

Review page 20

GRAPHS, CHARTS & FORMULAS FOR EM2

TRIGONOMETRY



Lengths of sides of right triangle: **$Hyp^2 = Op^2 + Adj^2$** or **$Hyp = (Op^2 + Adj^2)^{0.5}$**

Ratios: **Sine = Op/Hyp Cosine = Adj/Hyp Tangent = Op/Adj**

π = approx **3.14** 2π = approx **6.28** or **6.3** (close enough)

Circumference of a circle = **π x diameter** or **2π x radius**

RF BANDS

NAME	FREQ	WAVE-LENGTH	NAME	FREQ	WAVE-LENGTH
VLF very low frequency	3kHz - 30kHz	10kM - 100kM	VHF very high frequency	30MHz - 300MHz	1M - 10M
LF low frequency	30kHz - 300kHz	1kM - 10kM	UHF ultra high frequency	300MHz - 3GHZ	10CM - 1M
MF medium frequency	300kHz - 3MHz	100M-1kM	SHF super high frequency	3GHz - 30GHz	1CM - 10CM
HF high frequency	3MHz - 30MHz	10M - 100M	EHF extremely high frequency	30GHz - 300GHz	1MM - 1CM

SINE WAVE VOLTAGE AND POWER CONVERSIONS

$V_{RMS} = 0.707\ V_{PK}$ **$V_{PK} = 1.4\ V_{RMS}$**

$P_{AVE} = 0.5\ P_{PK}$ **$P_{PK} = 2\ P_{AVE}$**

CAPACITANCE AND INDUCTANCE FORMULAS

$X_C = 10^6 / 2\pi fC = 10^6 / 6.3fC$ where C is in microfarads

$X_L = 2\pi fL = 6.3fL$ where L is in farads

Review page 21

For capacitance: $dV / dT = I/C$

For inductance: $dI / dT = V/L$

For inductance or capacitance: $T = RL$ or $T = RC$

Time to full charge for a capacitor or full current for an inductor: 5 time constants

Series RC or RL: $Z = (R^2 + X^2)^{0.5}$ $V_T = (V_R^2 + V_C^2)^{0.5}$ or $(V_R^2 + V_L^2)^{0.5}$

Parallel RC or RL: $Z = (R \times X) / (R + X)$ $I_T = (I_R^2 + I_C^2)^{0.5}$ or $(I_R^2 + I_L^2)^{0.5}$

LC or RLC: $f = \frac{1}{2\pi\sqrt{LC}}$ or $f = \frac{0.159}{\sqrt{LC}}$

Q = X / R where X is either X_C or X_L at f

D = 1 / Q

TRANSFORMER IMPEDANCE

$$Z_{IN} / Z_{OUT} = (V_{IN} / V_{OUT})^2$$

MINIMUM LENGTH FOR A TRANSMISSION LINE

$L > 3 \times 10^7 / f$ length in meters
$L > 10^8 / f$ length in feet

RF CABLES

TYPE	*Z*	*DIAMETER*	*LOSS*
RG-6	**75Ω**	**medium**	**medium**
RG-8	**50Ω**	**medium**	**medium**
RG-11	**75Ω**	**large**	**low**
RG-58	**50Ω**	**small**	**high**
RG-59	**75Ω**	**small**	**high**
TWIN	**300Ω**		**low**

Review page 22

PAIRED WIRES COLOR CODE (CAT5 ETC)

pair/ wire	color	stripe	pair/ wire	color	stripe	pair/ wire	color	stripe	pair/ wire	color	stripe	pair/ wire	color	stripe
1+	WHT	BLU	**6+**	RED	BLU	**11+**	BLK	BLU	**16+**	YEL	BLU	**21+**	VIO	BLU
1-	BLU	WHT	**6-**	BLU	RED	**11-**	BLU	BLK	**16-**	BLU	YEL	**21-**	BLU	VIO
2+	WHT	OR	**7+**	RED	OR	**12+**	BLK	OR	**17+**	YEL	OR	**22+**	VIO	OR
2-	OR	WHT	**7-**	OR	RED	**12-**	OR	BLK	**17-**	OR	YEL	**22-**	OR	VIO
3+	WHT	GRN	**8+**	RED	GRN	**13+**	BLK	GRN	**18+**	YEL	GRN	**23+**	VIO	GRN
3-	GRN	WHT	**8-**	GRN	RED	**13-**	GRN	BLK	**18-**	GRN	YEL	**23-**	GRN	VIO
4+	WHT	BRN	**9+**	RED	BRN	**14+**	BLK	BRN	**19+**	YEL	BRN	**24+**	VIO	BRN
4-	BRN	WHT	**9-**	BRN	RED	**14-**	BRN	BLK	**19-**	BRN	YEL	**24-**	BRN	VIO
5+	WHT	GRY	**10+**	RED	GRY	**15+**	BLK	GRY	**20+**	YEL	GRY	**25+**	VIO	GRY
5-	GRY	WHT	**10-**	GRY	RED	**15-**	GRY	BLK	**20-**	GRY	YEL	**25-**	GRY	VIO

End of EM2 Review

EM2 PRACTICE EXAMS

Print these out and put them in a folder. Take the exams one chapter at a time, after you've read through the chapter and the review enough to be confident that you know the material. Answers for all exams are listed together, after the last exam.

The number of questions varies by chapter. ETA exams have a set format of 75 to 100 questions, but the purpose of these practice exams is not to mimic the ETA tests, but to allow you to find out how well you've learned the material. Therefore, there are as many questions for a particular chapter as it takes to cover each topic. There is no time limit when taking these exams, but you should probably keep track of your time so you'll have an idea how you will do when you take the actual ETA certification exams, which are timed. You may use a calculator, scratch paper and the charts and memory aids printed in the review.

CHAPTER 1 – INTRODUCTION

1. In alternating current:
    A. current flow alternates back and forth
    B. voltage alternates positive and negative
    C. resistance changes between high and low
    D. A and B above

2..Historically, DC came before AC because:
    A. DC is more useful
    B. DC is easier to transmit over long distances
    C. AC requires more advanced technology
    D. AC is less efficient

3. An AC voltage source schematic symbol:
    A. is a sine wave in a circle
    B. represents an idealized voltage source
    C. is considered to have zero internal resistance
    D. all of the above

4. The opposition to current flow presented by capacitors and inductors is called:
    A. Resistance
    B. Reactance
    C. Impedance
    D. Reluctance

5. Impedance is:
    A. any combination of resistance and reactance
    B. another name for resistance
    C. another name for reactance
    D. measured in watt-hours

CHAPTER 2 – MATH

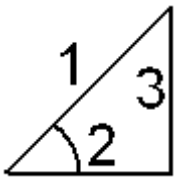
1. AC voltage, current, resistance and power are often indicated by:
   A. lower-case letters
   B. smaller-type letters
   C. Greek symbols
   D. letters in parentheses

2. The quadrants in a 4-quadrant X-Y graph are:
   A. 1, 2, 3, 4, clockwise from upper right
   B. A, B, C, D, clockwise from upper left
   C. I, II, III, IV, counter-clockwise from upper left
   D. I, II, III, IV, counter-clockwise from upper right

3. A phasor is:
   A. an arrow with a specific magnitude and direction
   B. often used to represent voltage or current
   C. considered to be rotating counter-clockwise
   D. all of the above

4. In a right triangle:
   A. all the internal angles add up to 90^{O}
   B. one internal angle is 90^{O}
   C. all the internal angles are the same
   D. every internal angle is 90^{O}

5. In the illustration at right, the Opposite, Adjacent and Hypotenuse are, respectively:



   A. 3,2,1
   B. 3,1,2
   C. 2,3,1
   D. 2,1,3

6. The sine, cosine and tangent of a right triangle are, respectively:
   A. Op / Adj, Adj / Op, Hyp / Adj
   B. Op / Hyp, Adj / Hyp, Op / Adj
   C. Adj / Op, Adj / Hyp, Op / Hyp
   D. Hyp / Op, Hyp/ Adj, Op / Adj

7. The Pythagorean theorem is:
   A. $\text{Hyp} = \sqrt{\text{Op}} + \sqrt{\text{Adj}}$
   B. $\text{Hyp} = \text{Op} + \text{Adj}$
   C. $\text{Hyp} = \sqrt{\text{Op}^2 + \text{Adj}^2}$
   D. $\sqrt{\text{Hyp}} = \text{Op} + \text{Adj}$

8. π is approximately:
   A. 0.707
   B. 6.3
   C. 3.14
   D. 3.14^2

9. The circumference of a circle is:
   A. π x radius squared
   B. π x radius
   C. π x diameter
   D. π / diameter

10. A radian:
   A. is the same length as a radius
   B. if multiplied by 2π, equal to circumference
   C. the same length as a diameter
   D. A and B above

11. Angular velocity is:
   A. the speed of a phasor tip
   B. the Greek letter ω
   C. $1 / 2\pi$ times frequency
   D. all of the above

12. An imaginary number corresponds to:
   A. a point on the X-axis to the left of the origin
   B. a point in quadrant III
   C. a point on the Y-axis
   D. a point in quadrant IV

13. Translate angular coordinates 3, 45^{O} to X,Y coordinates. Truncate to two places.
   A. 0.71, j0.71
   B. 2.12, j2.12
   C. 4.2, j4.2
   D. 4.2, -j4.2

14. A 3-cycle log scale might cover the frequency range of:
   A. 25kHz to 25MHz
   B. 100Hz to 300Hz
   C. 10^4 Hz to 10^8 Hz
   D. 3Hz t 30Hz

15. On a log scale, the physical space between 1 and 1.5 is the same as the space between:
   A. 2 and 4.5
   B. 3 and 6
   C. 4 and 6.5
   D. 5 and 7.5

CHAPTER 3 – WAVEFORMS

1. A periodic waveform is:
    A. one which has no distortion
    B. one which continuously repeats
    C. one with symmetrical + and - alternations
    D. a wave that turns on and off e.g. telegraph

2. The period of a wave is:
    A. $1 / \omega$
    B. frequency / speed
    C. 1 / f
    D. ω x f

3. The VHF band denotes the frequency range:
    A. 3MHz to 30MHz
    B. 30MHz to 300MHz
    C. 100 meters to 1kM
    D. 10 MHz to 100MHz

4. The speed of an electromagnetic signal in space is:
    A. 186,000 meters / second
    B. 300,000 miles / second
    C. 3×10^8 miles / second
    D. 3×10^8 meters / second

5. Wavelength is:
    A. speed / frequency
    B. frequency / speed
    C. angular velocity / frequency
    D. angular velocity / speed

6. The positive peak of a sine wave corresponds to:
    A. 90 degrees
    B. 45 degrees
    C. 180 degrees
    D. zero degrees

7. A complex wave is the combination of:
    A. square waves at various frequencies
    B. sawtooth waves at various frequencies
    C. sine waves at various frequencies
    D. noise and transients

8. Noise contains energy at:
    A. every frequency
    B. integer multiples of the fundamental frequency
    C. the fundamental frequency and DC
    D. only radio frequencies

9. Fast waveform rise and fall times are associated with:
    A. strong high-order harmonics
    B. radiation of interference
    C. inefficient transfer of power
    D. A and B above

10. A ramp is:
    A. a linear rise or fall in voltage over time
    B. a signal that gradually rises in amplitude
    C. a waveform that gradually rises in frequency
    D. a waveform that gradually rises in harmonic content

11. The RMS voltage of a sine wave is the same as the average voltage:
    A. true
    B. false

12. An inexpensive AC voltmeter gives an accurate AC voltage reading on the following:
    A. a 60Hz square wave
    B. a 20kHz sine wave
    C. a 60Hz sine wave
    D. any waveform with a frequency of 60Hz

13. The peak voltage of a sine wave is:
    A. 1.4 times the RMS voltage
    B. 1.4 times the absolute voltage
    C. 1.4 times the average voltage
    D. 0.707 times the RMS voltage

14. The peak power of a sine wave is:
    A. 1.4 times the average power
    C. twice the average power
    B. ½ the average power
    D. 0.707 times the average power

CHAPTER 4 – TEST EQUIPMENT & MEASUREMENTS

1. An oscilloscope is most commonly used to display:
    A. waveform distortion
    B. the spectrum of a waveform
    C. a signal waveform in real time
    D. phase delay of an RC circuit

2. The vertical size of the waveform on the display is adjusted using:
    A. the 5X setting
    B. the trigger threshold knob
    C. the vertical sensitivity switch
    D. the vertical position pot

3. “Triggering” refers to:
    A. synchronizing the sweep to the waveform
    B. vertically centering the sweep
    C. horizontally centering the sweep
    D. adjusting the horizontal sweep speed

4. The time interval displayed in a single sweep is adjusted using:
    A. the astigmatism control
    B. the horizontal sweep time switch
    C. the horizontal position control
    D. the trigger threshold control

5. The maximum voltage that can be displayed can be increased by:
    A. using a 10X attenuator probe
    B. using the X-Y setting
    C. manually setting the trigger voltage
    D. buying a better scope

6. Input loading refers to:
    A. the ratio of probe resistance to scope input resistance in 10X mode
    B. adjusting the input circuit so it is purely resistive
    C. selecting which input activates the trigger
    D. the load placed on the circuit under test by the oscilloscope

7. Setting the input switch to “AC”:
    A. adjusts the scope to run on 117VAC power
    B. causes the scope to display only DC, filtering out any AC at the input
    C. causes the horizontal sweep circuit to trigger on an AC voltage
    D. Inserts a capacitor in series with the vertical sensitivity switch

8. The X-Y setting of the sweep time switch:
    A. exchanges X and Y axes, so the waveform appears vertical
    B. displays input 1 vs. input 2 as an X-Y graph
    C. locks horizontal and vertical sensitivity controls so they are always equal
    D. switches the trigger input from input 1 to input 2

9. The best way to connect a scope ground to the ground of the unit under test is:
    A. via the clip that attaches to the probe
    B. via a separate wire to the scope chassis
    C. via the ground pins of the AC plugs of the scope and the unit under test
    D. the scope and the unit under test do not need to be grounded together.

10. An active probe:
    A. contains an amplifier in the probe
    B. reduces input loading on the circuit under test
    C. can easily be damaged
    D. all of the above

11. If a 10kHz square wave displayed on a scope has peaked or rounded corners,
    A. the scope or probe does not have flat response
    B. the vertical amplifier is overloading
    C. the vertical sensitivity knob is set wrong
    D. the scope's calibration output is defective

12. A storage oscilloscope:
    A. captures a waveform for later inspection
    B. averages successive waves
    C. can accommodate unusually high signal voltages
    D. is designed to display very low frequency waveforms

13. A spectrum analyzer displays:
    A. frequency on the Y-axis vs. amplitude on the X-axis
    B. frequency on the X-axis vs. amplitude on the Y-axis
    C. waveform distortion vs. time
    D. waveform change vs. time

14. A distortion analyzer usually displays:
    A. noise as a percentage of the signal
    B. distortion, measured in fractions of a volt
    C. distortion, measured in milliwatts
    D. distortion as a percentage of the signal

15. An LC tester displays:
    A. capacitance and inductance
    B. capacitance, inductance, partial shorts and opens
    C. resonant frequency of a capacitor or inductor
    D. breakdown voltage of a capacitor or inductor

16. A clamp-around ammeter displays:
    A. DC current flowing through one wire
    B. DC current flowing through a pair of wires
    C. AC current flowing through one wire
    D. AC current flowing through a pair of wires

17. Which is *not* a virtue of a function generator:
    A. output voltage stability
    B. ability to generate multiple waveforms
    C. low sine wave distortion
    D. ability to synchronize an oscilloscope

18. White noise contains equal energy in the frequency ranges:
    A. 300Hz – 700Hz and 1.5kHz – 1.9kHz
    B. 200Hz – 400Hz and 2kHz - 4kHz
    C. 1kHz – 3kHz and 1MHz - 3MHz
    D. 6kHz – 10kHz and 60kHz-60.1kHz

CHAPTER 5 – COMPONENTS

1. A capacitor stores energy in the form of:
    A. heat
    B. an electric charge across conductive plates
    C. an electric charge across conductive plates and in the insulating dielectric between them
    D. a flowing current

2. Capacitance is primarily determined by;
    A. plate size, distance between them, and type of dielectric
    B. plate size, distance between them, and their conductivity
    C. physical size of the capacitor
    D. breakdown voltage of the dielectric

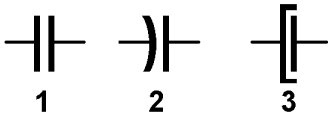
3. Proper DC polarity must be observed when using:
    A. a paper-dielectric capacitor
    B. an electrolytic capacitor
    C. a mica capacitor
    D. a ceramic capacitor

4. Moving the plates of a capacitor will result in:
    A. increased capacitance, increased breakdown voltage
    B. increased capacitance, decreased breakdown voltage
    C. decreased capacitance, increased breakdown voltage
    D. distance between plates does not matter, as long as they don't touch

5. In an axial-lead capacitor:
    A. one lead comes out each end
    B. both leads come out the same end
    C. the leads come out the side of the capacitor
    D. the leads are pre-bent for insertion into a PC board

6. The symbol for an electrolytic capacitor is:

    A. 2
    B. 1
    C. 1 or 2
    D. 2 or 3



7. A trimmer capacitor is:
    A. used to store large amounts of energy
    B. frequently used as the tuning means in a radio
    C. polarized
    D. a variable cap intended for infrequent adjustment

8. The name and symbol for a capacitor's opposition to current flow are:
    A. capacitive reactance, X_C
    B. capacitive reluctance, R_C
    C. capacitive impedance, Z_C
    D. capacitive inductance, L_C

9. AC current through a capacitor is out of step with voltage across it.
    A. false
    B. true
    C. depends on the circuit it's connected into
    D. depends on the frequency

10. An inductor coil stores energy in:
    A. the wire
    B. the wire's dielectric insulation
    C. the magnetic field around the coil
    D. an inductor does not store energy

11. An inductor's magnetic field only exists if:
    A. the coil is the right shape
    B. the right core material is used
    C. current is flowing through the coil
    D. there are no other inductors nearby

12. Which core material is equivalent to an air core?
    A. aluminum
    B. plastic
    C. ferrite
    D. carbon

13. Which core material/s can suffer from saturation?
    A. soft iron
    B. ceramic
    C. ferrite
    D. A and C

14. Laminations are used to:
    A. reduce the cost of construction
    B. reduce eddy currents
    C. reduce size and weight
    D. improve low frequency response

15. The name and symbol for an inductor's opposition to current flow are:
    A. inductive reluctance, X_L
    B. inductive reactance, R_L
    C. inductive impedance, Z_L
    D. inductive wavelength, ω_L

16. AC current through an inductor is out of step with the voltage across it.
    A. false
    B. true
    C. depends on the circuit it's connected into
    D. only at radio frequencies

17. When two inductors share a single field to some extent, this is called:
    A. reactive coupling
    B. interdependence
    C. mutual reactance
    D. mutual inductance

18. The coils of wire in a transformer are called:
    A. turns
    B. cores
    C. windings
    D. taps

19. A transformer whose output voltage is higher than its input voltage is called a_____ transformer
    A. primary
    B. secondary
    C. tuned
    D. step-up

20. Mutual inductance can be reduced by placing coils:
    A. at right angles to each other
    B. side by side
    C. as close as possible to each other
    D. within a common shielded enclosure

CHAPTER 6 – PRINCIPLES OF CAPACITANCE

1. The formula that describes the voltage/current relationship in a capacitor is:
    A. dT/dC = V/I
    B. dI/dT = V/C
    C. dV/dT = C/I
    D. dV/dT = I/C

2. The time constant formula is:
   A. $T = R / C$
   B. $T = RC$
   C. $T = RC^2$
   D. $T = C / X_C$

3. A capacitor connected so as to block DC but pass AC is called a _____ or _____ capacitor
   A. coupling, bypass
   B. bypass, shunt
   C. blocking, filter
   D. coupling, blocking

4. The formula for capacitive reactance is:
   A. $X_C = \frac{\pi f C}{2}$
   B. $X_C = 2\pi f C$
   C. $X_C = \frac{2}{\pi f C}$
   D. $X_C = \frac{1}{2\pi f C}$

5. When capacitors are connected in parallel:
   A. capacitance and X_C add
   B. capacitance adds, X_C falls
   C. X_C adds, capacitance falls
   D. Capacitance adds, X_C multiplies

6. When an AC sine wave is connected across a capacitor:
   A. current leads voltage by 90^O
   B. voltage leads current by 90^O
   C. current falls with increasing frequency
   D. phase angle depends on the circuit

7. When a capacitor is charged through a resistor, it is considered fully charged in:
   A. one time constant
   B. three time constants
   C. five time constants
   D. ten time constants

8. X_C of a 1μfd capacitor at 3kHz is:
   A. 53Ω
   B. 33Ω
   C. 300Ω
   D. 18.9Ω

9. If a capacitor's X_C is 9Ω at 10kHz, at 20kHz it will be:
   A. 18Ω
   B. 4.5Ω
   C. 12.6Ω
   D. 57Ω

10. If the X_C of an 0.01mfd capacitor is 4kΩ at a certain frequency, X_C of an 0.02mfd capacitor at the same frequency will be:
   A. 8kiΩ
   B. 5.6kΩ
   C. 2kΩ
   D. 2.8kΩ

CHAPTER 7 – PRINCIPLES OF INDUCTANCE

1. The formula that describes the voltage/current relationship in an inductor is:
   A. dT/dI = V/L
   B. dI/dT = V/L
   C. dV/dT = I/L
   D. dL/dT = V/I

2. The time constant formula is:
   A. T = RL
   B. $T = RL^2$
   C. T = R / L
   D. T = R + L

3. The property that keeps current from rising instantly when voltage is applied is:
   A. back emf
   B. residual magnetism
   C. mutual inductance
   D. mutual resistance

4. A flyback is:
   A. a magnetized inductor's tendency to attract iron objects
   B. physical vibration of a coil due to self-induction
   C. a voltage in an inductor coil induced by an external magnetic field
   D. a voltage transient generated when a current-carrying inductor is disconnected

5. The formula for inductive reactance is:

   A. $X_L = \frac{\pi f}{2L}$ C. $X_L = \frac{2}{\pi f L}$

   B. $X_L = 2\pi f L$ D. $X_L = \frac{1}{2\pi f L}$

6. When inductors are connected in parallel:
    A. inductance rises, X_L falls
    B. X_L rises, inductance falls
    C. inductance and X_L both fall
    D. inductance and X_L both rise

7. When an AC sine wave is connected across an inductor:
    A. voltage lags current by 90
    B. saturation lags voltage by 90^O
    C. the magnetic field and current are 90^O out of phase
    D. current lags voltage by 90^O

8. When an inductor is connected to a DC voltage source through a resistor, current reaches essentially its final value in:
    A. current continues to rise forever
    B. one time constant
    C. three time constants
    D. five time constants

9. X_L of a 1mH inductor at 2kHz is:
    A. $12.6 \times 10^6 W$
    B. 12.6kW
    C. 12.6Ω
    D. 8Ω

10. If an inductor has an X_L of 50Ω at a 1kHz, X_L at 2kHz will be:
    A. 25Ω
    B. 35Ω
    C. 100Ω
    D. 71Ω

CHAPTER 8 – TRANSFORMER THEORY

1. The step up/step down ratio is determined by:
    A. the ratio of wire gauge in the primary to wire gauge in the secondary
    B. the turns ratio
    C. the impedance ratio
    D. the ratio of the diameters of the coils

2. In a 5:1 transformer, if the load on the secondary is 3A, what is the load on the voltage source?
    A. 15A
    B. 0.6A
    C. 1A
    D. 0.12A

3. If the voltage across a secondary is 18V and it has a tap 1/3 of the way from one end of the winding, voltage between the tap and that end will be:

A. 6V
B. 9V
C. 1.7V
D. 3V

4. A particular transformer has a primary with 300 turns and two secondaries with 100 and 30 turns respectively. If the primary is connected to 117VAC and the secondaries are connected in series, what will the output voltage be?

A. 270V
B. 130V
C. 49V
D. 51V

5. The advantage of a toroid power transformer are:

A. smaller size for a given power level
B. less stray magnetic radiaton
C. lower cost
D. A and B above

6. The low-frequency limit of an iron-core transformer is set by:

A. number of turns in primary coil
B. wire gauge of primary coil
C. core saturation
D. degree of mutual inductance

7. The high-frequency limit of an iron-core transformer is set by:

A. number of turns in secondary coil
B. weight of iron in the core
C. dielectric constant of wire insulation
D. stray inductance and core losses

8. The impedance ratio of a 3:1 stepdown transformer is:

A. 9:1
B. 3:1
C. 1.7:1
D. 6:1

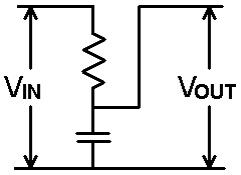
9. The reduction in output voltage that occurs as the load current is increased is called:

A. reluctance
B. impedance
C. step down
D. sag

10. The voltage drop across the primary of a current transformer in a 117V circuit is:
    A. equal to the voltage across the secondary
    B. 117V
    C. very small
    D. depends on the loading of the secondary

CHAPTER 9 – RC, RL, LC, RLC PRINCIPLES

Questions 1-4 refer to the circuit at right.



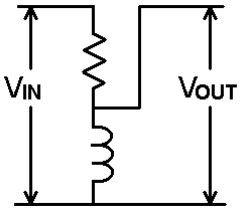
1. The sum of the voltage drops across R and C will be _____ the input voltage.
    A. less than
    B. equal to
    C. opposite polarity from
    D. greater than

2. If R = X_C and V_{IN} = 1V, V_{OUT} will be:
    A. 0.707V
    B. 1.4V
    C. 0.5V
    D. 0.36V

3. If R = X_C, the phase of V_{OUT} will be _____ relative to V_{IN}.
    A. -45^O
    B. -90^O
    C. 90^O
    D. depends on input voltage

4. If the frequency is raised, what will happen to the output?
    A. voltage and phase lag will increase
    B. voltage will decrease, phase lead will increase
    C. voltage will decrease, phase lag will increase
    D. voltage and phase lead will increase

Questions 5-7 refer to the figure at right.



5. If R = X_L and V_{IN} is 1.4V, V_{OUT} will be:

    A. 1V
    B. 0.707V
    C. 2V
    D. 0.5V

6. If R = X_L, the phase of V_{OUT} will be _____ relative to V_{IN}:
   A. -45^O
   B. -90^O
   C. 90^O
   D. 45^O

7. If the frequency is raised, what will happen to the output?
   A. voltage and phase lead will increase
   B. voltage will increase, phase lead will decrease
   C. voltage and phase lag will increase
   D. voltage will decrease, phase lag will increase

8. The formula for impedance of a series RC or RL circuit is:
   A. $Z = \sqrt{R} + \sqrt{X}$
   B. $Z = R^2 + X^2$
   C. $Z = \dfrac{1}{\sqrt{R^2 + X^2}}$
   D. $Z = \sqrt{R^2 + X^2}$

9. The formula for current in a series RC or RL circuit is:
   A. I = VZ
   B. I = Z / V
   C. $I = V^2 / Z$
   D. I = V / Z

10. The formula for current in a parallel RC circuit is:
   A. $I_T = \sqrt{I_R} + \sqrt{I_C}$
   B. $I_T = I_R \times I_C$
   C. $I_T = \sqrt{I_R^2 + I_C^2}$
   D. $I_T = \dfrac{1}{\sqrt{I_R^2 + I_C^2}}$

11. The formula for reactance in a series LC circuit is_____, whichever gives a positive result.
   A. $X = X_C / X_L$ or X_L / X_C
   B. $X = Z_C - Z_L$ or $Z_L - Z_C$
   C. $X = X_C - X_L$ or $X_L - X_C$
   D. $X = 1 / (X_C - X_L)$ or $1 / (X_L - X_C)$

12. At frequency f, in a particular series LC circuit $X_C = 100\Omega$, $X_L = 100\Omega$ and there is 1Ω of resistance. Z is:

A. 1Ω
B. 100Ω
C. 201Ω
D. 65Ω

13. In the circuit of question 12, at 2f the total reactance would be:

A. 150Ω, capacitive
B. 150Ω, reactive
C. 250Ω, reactive
D. 250Ω, capacitive

14. The formula for the resonant frequency of an LC circuit is: (two correct answers)

A. $f = 0.159\sqrt{LC}$

B. $f = \frac{0.159}{\sqrt{LC}}$

C. $f = 2\pi\sqrt{LC}$

D. $f = \frac{1}{2\pi\sqrt{LC}}$

15. A 1μfd capacitor and 1mH inductor resonate at approximately:

A. 160MHz
B. 200kHz
C. 5kHz
D. 2kHz

16. Changing the capacitor in question 15 to 2mfd would change the resonant frequency to approximately:

A. 3.5kHz
B. 1.4kHz
C. 280kHz
D. 2.5kHz

17. In a resonant circuit where $X_C = X_L = 100\Omega$ and $R = 10\Omega$, Q is:

A. 10
B. 20
C. 5
D. 7.07

18. In a circuit with a Q of 20, damping is:

A. -20
B. 20
C. 0.05
D. 1/5

19. Ideally, the source and load Z connected to most RC or RL filters are ____ and ____.
    A. 600Ω
    B. equal to X_C or X_L
    C. equal to R
    D. zero, infinity

20. Proper operation of a π or T filter depends on:
    A. correct source impedance
    B. correct load impedance
    C. both
    D. source and load impedance are irrelevant

CHAPTER 10 – CABLING

1. Which is not a cable?
    A. a clip lead
    B. a 117V extension cord
    C. a satellite TV connection
    D. a telephone cord

2. A ground potential is:
    A. a voltage difference between two grounds
    B. a voltage difference between a unit ground and Earth ground
    C. a DC voltage between a signal connection and a ground connection
    D. what happens when a unit's power supply is shorted to ground

3. Noise, from a cabling standpoint, is:
    A. lamp dimmer buzz
    B. stray radio signals
    C. leakage from one signal wire to another
    D. all of the above

4. The two types of radiated noise are:
    A. electromagnetic and electrodynamic
    B. resistive and reactive
    C. electromagnetic and electrostatic
    D. AC and DC

5. Two factors that seriously limit how “hot” a signal can be are:
    A. the possibility of melting the cable
    B. the maximum output capability of an amplifier
    C. the potential for crosstalk
    D. B and C above

6. For a cable shield to be effective, it usually must be:
    A. braided
    B. grounded to the equipment on either end
    C. tightly wound around the center conductor
    D. made of aluminum foil with a copper drain wire

7. A cable's shield is effective against:
    A. electromagnetic radiation
    B. electromagnetic radiation
    C. both
    D. neither

8. Shielded cables should not be used for speaker wiring because:
    A. they heat up on loud music
    B. they have high resistive signal loss
    C. they have comparatively high inductance
    D. they have comparatively high capacitance

9. The signal on a balanced pair is:
    A. high level
    B. earth grounded
    C. differential mode
    D. common mode

10. The reason for twisting a balanced pair is:
    A. to improve the efficiency of signal transmission
    B. to reduce high frequency loss
    C. to make sure noise is common mode
    D. to make sure noise is differential mode

11. An unbalanced signal is:
    A. a single voltage measured against equipment ground
    B. a single voltage measured against Earth ground
    C. a defective signal
    D. two signals of unequal level

12. The main purpose of a balun is to:
    A. invert the polarity of a signal
    B. connect a balanced and unbalanced circuit together
    C. adjust the level of a signal
    D. change the impedance of a signal

13. A “floating” balanced pair refers to:
    A. a pair of wires that is transformer-isolated at both ends
    B. a pair of wires connected to a piece of equipment that is not earth-grounded
    C. a pair of wires where neither wire is connected to the shield
    D. a pair of wires that is not impedance-matched at either end

14. Plugging various pieces of equipment in a system into grounded outlets in different places is:
    A. required by safety codes
    B. a good way to improve the S/N ratio
    C. a good way to ruin the S/N ratio
    D. not likely to have any effect

15. In an impedance matched system:
    A. every input and output is 600 ohms
    B. every input and output is purely resistive
    C. power transfer efficiency is greatest
    D. every output is connected to a load equal to its own impedance

16. An RG-58 cable connected to a unit with a 50Ω input impedance is:
    A. mated
    B. joined
    C. spliced
    D. terminated

17. If an RG-59 cable is not terminated, when a signal reaches the end of the cable, it will be:
    A. reflected back toward the source
    B. absorbed into the atmosphere
    C. radiated into space
    D. dissipated as heat

18. True or false: a 30-foot coax cable carrying a 1MHz signal is a transmission line.
    A. true
    B. false

19. True or false: a 30-foot coax cable carrying a 100MHz signal is a transmission line.
    A. true
    B. false

20. Unshielded, balanced pairs have become popular at very high frequencies because:
    A. they are less expensive than shielded cables
    B. they are physically smaller than shielded cables
    C. they have lower attenuation at high frequencies
    D. they are easier to work with

ANSWERS

Chapter 1, Basic Alternating Current

1-D 2-C 3-D 4-B 5-A

Chapter 2, AC Math

1-A 2-D 3-D 4-B 5-A 6-B 7-C 8-C 9-C 10-D 11-D 12-C 13-B 14-A 15-D

Chapter 3, Waves and Waveforms

1-A 2-D 3-A 4-B 5-A 6-B 7-C 8-C 9-C 10-D 11-D 12-C 13-B 14-A15-D

Chapter 4, Test Equipment and Measurements

1-C 2-C 3-A 4-B 5-A 6-D 7-D 8-B 9-B 10-D 11-A 12-A 13-B 14-D 15-B 16-C 17-C 18-A

Chapter 5, Components

1-C 2-A 3-B 4-B 5-B 6-D 7-D 8-A 9-B 10-C 11-C 12-B 13-D 14-B 15-B 16-B 17-D 18-C 19-D 20-A

Chapter 6, Principles of Capacitance

1-D 2-B 3-D 4-D 5-B 6-A 7-C 8-A 9-B 10-C

Chapter 7, Principles of Inductance

1-B 2-A 3-A 4-D 5-B 6-C 7-D 8-D 9-C 10-C

Chapter 8, Transformer Theory

1-B 2-B 3-A 4-C 5-D 6-C 7-D 8-A 9-D 10-C

Chapter 9, RC, RL, LC, RLC Principles

1-D 2-A 3-A 4-C 5-A 6-D 7-B 8-D 9-D 10-C 11-C 12-A 13-B 14-B,D 15-C 16-A 17-A 18-C 19-D 20-C

Chapter 10, Cabling

1-A 2-A 3-D 4-C 5-A 6-B 7-D 8-B 9-C 10-C 11-A 12-B 13-A 14-C 15-D 16-D 17-A 18-B 19-A 20-C

End of EM2 Practice Exams